Wahrscheinlichkeitsrechnung und Statistik für Biologen
3. Grundlagen aus der Wahrscheinlichkeitstheorie

Dirk Metzler

28. April 2021

Inhaltsverzeichnis

1 Deterministische und zufällige Vorgänge

Nehmen wir an, in einer kleinen Population von $n = 100$ Individuen hat ein neutrales Allel A in der aktuellen Generation eine Häufigkeit von 0.32.

Wie wird die Häufigkeit X von A in der nächsten Generation sein?

Das können wir nicht genau vorhersagen, denn es hängt vom Zufall ab.

X ist eine Zufallsvariable.

Berechnen können wir aber z.B.:

$\mathbb{E}X$ **:** den Erwartungswert von X,

$\Pr(X = 0.32)$ **:** die Wahrscheinlichkeit, dass X einen bestimmten Wert annimmt.

Allerdings müssen wir dazu die Population zunächst stochastisch modellieren, und das Ergebnis wird davon abhängen, wie wir das machen.

Was können wir vorhersagen:

- Freier Fall: Falldauer eines Objektes bei gegebener Fallhöhe läßt sich vorhersagen (falls Luftwiderstand vernachlässigbar)

  Deterministische Vorgänge laufen immer gleich ab. Aus Beobachtungen lassen sich künftige Versuche vorhersagen.

Was können wir vorhersagen:

- Würfelwurf: Das Ergebnis eines einzelnen Würfelwurfes lässt sich nicht vorhersagen.
- Wiederholter Würfelwurf:

  Würfelt man 600 mal, so würde man gerne darauf wetten, dass die Anzahl an Einsern zwischen 75 und 125 liegt.

  Die genaue Anzahl lässt sich wieder nicht vorhersagen.

  Aber: Eine Aussage über die Verteilung ist möglich (die besser ist als reines Raten.)

Empirisch stellt man fest:

Bei Wiederholung eines Zufallsexperiments stabilisieren sich die relativen Häufigkeiten der möglichen Ergebnisse.

Beispiel:

Beim Würfelwurf stabilisiert sich die relative Häufigkeit jeder der Zahlen $\{1, 2, \ldots, 6\}$ bei $\frac{1}{6}$.

Fazit:

Das Ergebnis eines einzelnen zufälligen Vorgangs lässt sich nicht vorhersagen. Aber: Eine Aussage über die Verteilung ist möglich (die besser ist als reines Raten).

Abstraktionsschritt:

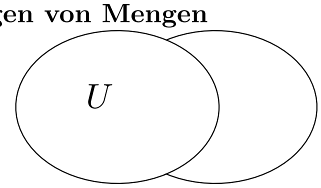
Verwende empirisch ermittelte Verteilung als Verteilung jedes Einzelexperiments!

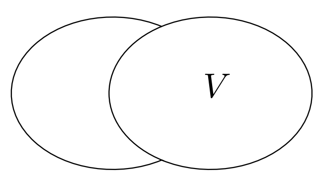
Beispiel:

Wir nehmen an, dass bei einem einzelnen Würfelwurf jede der Zahlen $\{1, 2, \ldots, 6\}$ die Wahrscheinlichkeit $\frac{1}{6}$ hat.

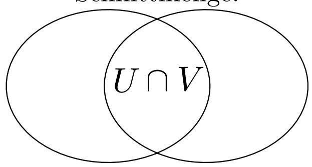
2 Zufallsvariablen und Verteilung

Verknüpfungen von Mengen



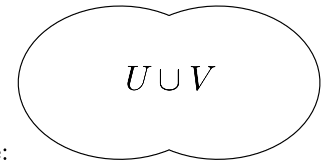


Schnittmenge:



$$U \cap V = \{x : x \in U \wedge x \in V\}$$

($\wedge$ ist “und”)



Vereinigungsmenge:

$$U \cup V = \{x : x \in U \vee x \in V\}$$

($\vee$ ist “oder”)

Ein Würfel werde geworfen, und die folgenden Ereignisse beziehen sich auf diesen einen Wurf:

$$\begin{aligned}
A &= \{\text{Das Ergebnis ist 2}\} \qquad \Pr(A) = \frac{1}{6} \\
B &= \{\text{Das Ergebnis ist nicht 5}\} \qquad \Pr(B) = 1 - \frac{1}{6} = \frac{5}{6} \\
C &= \{\text{Das Ergebnis ist kleiner als 4}\} \qquad \Pr(C) = \frac{3}{6} = \frac{1}{2} \\
D &= \{\text{Das Ergebnis ist gerade}\} \qquad \Pr(D) = \frac{3}{6} = \frac{1}{2}
\end{aligned}$$

Wir können Ereignisse mit “und” (Symbol $\cap$ oder $\wedge$) oder “oder” (Symbol $\cup$ oder $\vee$) verknüpfen:

$$\begin{aligned}
C \cap D &= C \wedge D = A \\
C \cup D &= C \vee D = B
\end{aligned}$$

Wir können die vorherigen Ereignisse kürzer und übersichtlicher schreiben, indem wir das Ergebnis des Würfelwurfs als Zufallsvariale W schreiben:

$$\begin{aligned}
A &= \{\text{Das Ergebnis ist 2}\} = \{W = 2\} \\
B &= \{\text{Das Ergebnis ist nicht 5}\} = \{W \neq 5\} = \{W \in \{1,2,3,4,6\}\} \\
C &= \{\text{Das Ergebnis ist kleiner als 4}\} = \{W < 4\} = \{W \in \{1,2,3\}\} \\
D &= \{\text{Das Ergebnis ist gerade}\} = \{W \in \{2,4,6\}\}
\end{aligned}$$

$$\begin{aligned}
C \wedge D &= \{W \in \{1,2,3\}\} \wedge \{W \in \{2,4,6\}\} = \{W \in \{1,2,3\} \cap \{2,4,6\}\} \\
&= \{W = 2\} = A \\
C \vee D &= \{W \in \{1,2,3\}\} \vee \{W \in \{2,4,6\}\} = \{W \in \{1,2,3\} \cup \{2,4,6\}\} \\
&= \{W = \{1,2,3,4,6\}\} = B
\end{aligned}$$

Als Zufallsgröße oder Zufallsvariable bezeichnet man das (Mess-)Ergebnis eines zufälligen Vorgangs.

Der Wertebereich $\mathcal{S}$ (engl. state space) einer Zufallsgröße ist die Menge aller möglichen Werte.

Die Verteilung einer Zufallsgröße X weist jeder Menge $A \subseteq \mathcal{S}$ die Wahrscheinlichkeit $\Pr(X \in A)$ zu, dass X einen Wert in A annimmt.

Für Zufallsgrößen werden üblicherweise Großbuchstaben verwendet (z.B. X, Y, Z), für konkrete Werte Kleinbuchstaben.

Mengenschreibweise

Das Ereignis, dass X einen Wert in A annimmt, kann man mit geschweiften Klammern schreiben:

$$\{X \in A\}$$

Dies kann man interpretieren als die Menge aller Elementarereignisse, für die X einen Wert in A annimmt. Die Schnittmenge

$$\{X \in A\} \cap \{X \in B\} = \{X \in A, X \in B\}$$

ist dann das Ereignis, dass der Wert von X in A liegt und in B liegt.
Die Vereinigungsmenge

$$\{X \in A\} \cup \{X \in B\}$$

ist das Ereignis, dass der Wert von X in A liegt oder (auch) in B liegt.
Bei Wahrscheinlichkeiten lässt man die Klammern oft weg:

$$\Pr(X \in A, X \in B) = \Pr(\{X \in A, X \in B\})$$

Beispiel: Würfelwurf W = Augenzahl des nächsten Würfelwurfs.

$\mathcal{S} = \{1, 2, \ldots, 6\}$ $\Pr(W = 1) = \cdots = \Pr(W = 6) = \frac{1}{6}$ ($\Pr(W = x) = \frac{1}{6}$ für alle $x \in \{1, \ldots, 6\}$) Die Verteilung erhält man aus einer Symmetrieüberlegung oder aus einer langen Würfelreihe.

Beispiel: Körpergrößenverteilung 20-jähriger Münchner.

Wähle aus dem Einwohnerregister einen zufälligen 20-jährigen. Sei K seine Körpergröße in cm. Dann ist K eine Zufallsvariable.

Die Verteilung von K ist nicht bekannt, lässt sich aber aus Stichproben schätzen.

Rechenregeln:

Beispiel Würfelwurf W:

$$\begin{aligned}
\Pr(W \in \{2, 3\}) &= \frac{2}{6} = \frac{1}{6} + \frac{1}{6} \\
&= \Pr(W = 2) + \Pr(W = 3) \\
\Pr(W \in \{1, 2\} \cup \{3, 4\}) &= \frac{4}{6} = \frac{2}{6} + \frac{2}{6} \\
&= \Pr(W \in \{1, 2\}) + \Pr(W \in \{3, 4\})
\end{aligned}$$

Vorsicht:

$$\begin{aligned}
\Pr(W \in \{2, 3\}) + \Pr(W \in \{3, 4\}) &= \frac{2}{6} + \frac{2}{6} = \frac{4}{6} \\
&\neq \Pr(W \in \{2, 3, 4\}) = \frac{3}{6}
\end{aligned}$$

Wir können auch Ereignissen Namen geben:

$$\begin{aligned}
U &:= \{X \in A\}, \qquad V := \{X \in B\} \\
\Rightarrow \quad U \cap V &= \{X \in A \cap B\}
\end{aligned}$$

Falls Ereignisse einander widersprechen, z.B.

$$U = \{X \in \{1, 2\}\} \quad V = \{X \in \{3, 4\}\}$$

dann

$$U \cap V = \emptyset = \{X \in \emptyset\},$$

wobei $\emptyset$ das unmögliche ("leere") Ereignis ist (das mit dem selben Symbol wie die leere Menge geschrieben wird).

Ereignisse sind Teilmengen einer Menge Ω von Elementarereignissen. Beispiel: ist X das Ergebnis eines Würfelwurfs, ist der Ergebnisraum

$$\Omega = \Big\{\{X = 1\}, \{X = 2\}, \{X = 3\}, \{X = 4\}, \{X = 5\}, \{X = 6\}\Big\}.$$

- Ist Ω endlich, wie hier, sind alle Teilmengen $U \subseteq \Omega$ Ereignisse und $\Pr(U) = \sum_{x \in U} \Pr(x)$.
- Für unendliche Ω kann es komplizierter werden:
  - $\Pr(U) > 0$ is möglich, auch wenn $\Pr(x) = 0$ für alle $x \in U$.
  - Es kann $U \subset \Omega$ geben, die nicht als Ereignisse zulässig sind (aus komplizierten mathematischen Gründen).
- Eine Wahrscheinlichkeitsverteilung ordnet jedem Ereignis $U \subseteq \Omega$ eine Wahrscheinlichkeit $\Pr(U)$ zu.

Elementare Regeln für Wahrscheinlichkeiten:

- $0 \leq \Pr(U) \leq 1$ für alle Ereignisse $U \subseteq \Omega$
- Ω und das leere Ereignis $\emptyset$ sind Ereignisse, und $\Pr(\Omega) = 1$ und $\Pr(\emptyset) = 0$.
- Ist U ein Ereignis, dann ist auch das Komplement (oder "Gegenereignis") $U^c = \Omega \setminus U$ ein Ereignis und es gilt
$$\Pr(U^c) = 1 - \Pr(U)$$
- Sind U und V Ereignisse, so sind auch auch $U \cup V$ und $U \cap V$ Ereignisse, und es gilt die Einschluss-Ausschluss-Formel:
$$\Pr(U \cup V) = \Pr(U) + \Pr(V) - \Pr(U \cap V)$$
(Falls also $U \cap V = \emptyset$, dann $\Pr(U \cup V) = \Pr(U) + \Pr(V)$.)

Beispiel zweifacher Würfelwurf (W_1, W_2): Sei W_1 (bzw W_2) die Augenzahl des ersten (bzw zweiten) Würfels.

$$\begin{aligned}
&\Pr(W_1 \in \{4\}, W_2 \in \{2, 3, 4\}) \\
&= \Pr((W_1, W_2) \in \{(4, 2), (4, 3), (4, 4)\}) \\
&= \frac{3}{36} = \frac{1}{6} \cdot \frac{3}{6} \\
&= \Pr(W_1 \in \{4\}) \cdot \Pr(W_2 \in \{2, 3, 4\})
\end{aligned}$$

Allgemein:

$$\Pr(W_1 \in A, W_2 \in B) = \Pr(W_1 \in A) \cdot \Pr(W_2 \in B)$$

für alle Mengen $A, B \subseteq \{1, 2, \ldots, 6\}$

Stochastische Unabhängigkeit

Definition 1 (stochastische Unabhängigkeit) *Ereignisse U und V sind sind (stochastisch) unabhängig, wenn gilt*

$$\Pr(U, V) = \Pr(U) \cdot \Pr(V).$$

Zwei Zufallsgrößen X und Y sind (stochastisch) unabhängig, wenn jedes Ereignis der Art $\{X \in A\}$ von jedem Ereignis der Art $\{Y \in B\}$ stochastisch unabhängig ist.

Beispiel:

- Werfen zweier Würfel: X = Augenzahl Würfel 1, Y = Augenzahl Würfel 2.

$$\Pr(X = 2,\ Y = 5) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \Pr(X = 2) \cdot \Pr(Y = 5)$$

"Stochastisch unabhängig" bedeutet nicht immer, das zwei Zufallsexperimente ganz unabhängig voneinander durchgeführt wurden. Z.B. beim Ziehen einer Karte aus einem Spiel mit 32 Karten sei

Ereignis A: Die gezogene Karte ist ein Ass

Ereignis $\heartsuit$: Die gezogene Karte ist von der Farbe Herz

Dann gilt

$$\Pr(\heartsuit, A) = \frac{1}{32} = \frac{1}{4} \cdot \frac{1}{8} = \Pr(\heartsuit) \cdot \Pr(A).$$

Also sind die Ereignisse A und $\heartsuit$ stochstisch unabhängig.

- Bedingte Wahrscheinlichkeit: Wahrscheinlichkeit des Ereignisses U unter der Bedingung V

$$\Pr(U|V) := \frac{\Pr(U, V)}{\Pr(V)}$$

„bedingte Wahrscheinlichkeit von U, gegeben V"

Daraus ergibt sich also auch:

$$\Pr(U, V) = \Pr(V) \cdot \Pr(U|V) = \Pr(U) \cdot \Pr(V|U)$$

Sei S die Summe der Augenzahlen, d.h. $S = W_1 + W_2$. Was ist die Wahrscheinlichkeit, dass $S = 5$ ist, wenn der erste Würfel die Augenzahl $W_1 = 2$ zeigt?

$$\Pr(S = 5|W_1 = 2) \overset{!}{=} \Pr(W_2 = 3)$$
$$= \tfrac{1}{6} = \tfrac{1/36}{1/6} = \tfrac{\Pr(S=5, W_1=2)}{\Pr(W_1=2)}$$

Was ist die Ws von $S \in \{4, 5\}$ unter der Bedingung $W_1 = 1$?

$$\begin{aligned}
&\Pr(S \in \{4, 5\}|W_1 = 1) \\
&\overset{!}{=} \Pr(W_2 \in \{3, 4\}) \\
&= \frac{2}{6} = \frac{2/36}{1/6} = \frac{\Pr(W_2 \in \{3, 4\}, W_1 = 1)}{\Pr(W_1 = 1)} \\
&= \frac{\Pr(S \in \{4, 5\}, W_1 = 1)}{\Pr(W_1 = 1)}
\end{aligned}$$

Nun mal umgekehrt: Was ist die Wahrscheinlichkeit, dass der erste Würfel eine 1 zeigt, wenn bekannt ist, dass die Summe 4 oder 5 ist?

$$\Pr(W_1 = 1 \mid S \in \{4, 5\}) =?$$

Zur Erfüllung der Bedingung $\{S \in \{4, 5\}\}$ gibt 7 Möglichkeiten für das Wertepaar (W_1, W_2):

$$(1, 3), (1, 4), (2, 2), (2, 3), (3, 1), (3, 2), (4, 1)$$

Davon sind nur folgenden beiden geeignet zur Erfüllung von $\{W_1 = 1\}$:

$$(1, 3), (1, 4).$$

Mit der Regel „Anzahl günstige durch Anzahl mögliche" erhalten wir also

$$\Pr(W_1 = 1 \mid S \in \{4, 5\}) = \frac{2}{7} = \frac{\frac{2}{36}}{\frac{7}{36}} = \frac{\Pr(W_1 = 1, S \in \{4, 5\})}{\Pr(S \in \{4, 5\})}.$$

Rechnen mit Zufallsvariablen

Am Beispiel der Summe $W_1 + W_2$ zweier Würfelergebnisse W_1, W_2 haben wir gesehen, dass man mit Zufallsvariablen auch wie mit normalen Variablen rechnen kann. Auch $W_1 + W_2$ ist wieder eine Zufallsvariable und es gilt:

$$\begin{aligned}\Pr(W_1 + W_2 = k) &= \sum_{\{(x,y)\ :\ x+y=k\}} \Pr(W_1 = x,\ W_2 = y) \\ &= \sum_x \Pr(W_1 = x,\ W_2 = k - x)\end{aligned}$$

In Worten: $\Pr(W_1 + W_2 = k)$ ist die Summe der Wahrscheinlichkeiten aller Wertepaare für W_1 und W_2, die in Summe k ergeben.

Rechnen mit Zufallsvariablen

Sie können Zufallsvariablen auch mit Zahlen multiplizieren. Stellen Sie sich z.B. vor, dass Sie bei einem Brettspiel die "Speed-Karte" gezogen haben, die Ihnen erlaubt, immer das Doppelte $D = 2 \cdot W$ der gewürfelten Augenzahl W voranzuschreiten.

Dann ist $D = 2 \cdot W$ ebenfalls eine Zufallsvariable. Ihre möglichen Werte sind 2,4,6,8,10 und 12, und D nimmt jeden dieser Werte mit Wahrscheinlichkeit $\frac{1}{6}$ an.

Anwendung einer Funktion auf eine Zufallsvariable

Wenn man eine Funktion f auf eine Zufallsvariable X anwendet, erhält man eine Zufallsvariable $Y = f(X)$.

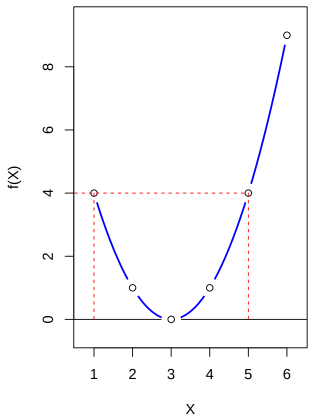
Als Beispiel stellen wir uns ein Spiel vor, bei dem man, wenn man das Würfelergebnis X hat, $f(X) = (X - 3)^2$ Felder vorrücken darf.

Dann ist die Wahrscheinlichkeit, dass man 4 Felder vorrücken darf, also dass die Zufallsvariable $f(X)$ den Wert 4 hat (also X den Wert 1 oder 5):

$$\Pr(f(X) = 4) = \Pr(X \in \{1, 5\}) = 1/3.$$

Um das mathmatisch klar zu fassen, benötigen wir den Begriff **Urbild**: Für eine Menge M ist $f^{-1}(M) = \{x\ :\ f(x) \in M\}$ das **Urbild** von M unter f, also in Worten: die Menge aller x, die von f auf ein Element von M abgebildet werden.

Das Urbild einer Menge unter einer Funktion f

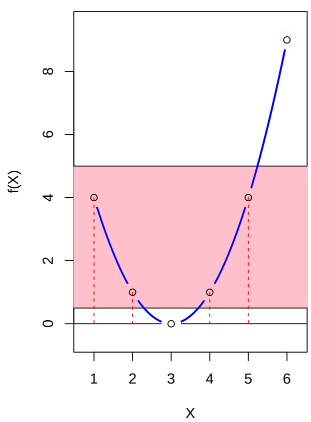


Die Funktion $f : x \mapsto (x-3)^2$ für $x \in \{1,2,3,4,5,6\})$ ist nicht umkehrbar. So ist $f^{-1}(4)$ nicht definierbar, denn es wäre ja auch $f(1) = 4 = f(5)$. Etwas anderes ist jedoch $f^{-1}(\{4\})$. Hier ist mit f^{-1} keine Umkehrfunktion gemeint, sondern die
Urbildfunktion, die auf Mengen operiert:

$$f^{-1}(\{4\}) \;=\; \{x \;:\; f(x) \in \{4\}\} \;=\; \{1,5\}$$

Oder z.B.:

$$f^{-1}([0.5,5]) \;=\; \{x \;:\; f(x) \in [0.5,5]\} \;=\; \{1,2,4,5\}$$



Also ist die Wahrscheinlichkeit, dass man 4 Felder vorrücken darf, also dass die Zufallsvariable $f(X)$ den Wert 4 hat (also X den Wert 1 oder 5):

$$\Pr(f(X) = 4) = \Pr(X \in f^{-1}(\{4\})) = \Pr(X \in \{1,5\}) = 1/3$$

Entsprechend gilt:

$$\Pr(f(X) \in [0.5,5]) = \Pr(X \in f^{-1}([0.5,5])) = \Pr(X \in \{1,2,4,5\}) = 2/3$$

3 Die Binomialverteilung

0-1-Zufallsvariablen

Zufallsvariablen, die nur die Werte 0 oder 1 annehmen können, bezeichnet man als

0-1-Zufallsvariablen

oder auch als Bernoulli-verteilte Zufallsvariablen.

Sei A ein Ereignis, z.B. {Münzwurf zeigt Kopf} oder {Drosophila hat Mutation für weiße Augen}. Dann ist die I_A die Indikatorvariable für A die 0-1-Zufallsvariable mit:

$$I_A = \begin{cases} 1 & \text{falls } A \text{ eintritt} \\ 0 & \text{falls } A \text{ nicht eintritt} \end{cases}$$

Der Parameter $p = \Pr(I_A = 1) = \Pr(A)$ heißt Erfolgswahrscheinlichkeit.

Angenommen, ein Experiment (z.B. Münzwurf zeigt Kopf) mit Erfolgswahrscheinlichkeit p, wird n mal *unabhängig* wiederholt.

Wie groß ist die Wahrscheinlichkeit, dass es...

1. ...immer gelingt?
$$p \cdot p \cdot p \cdots p = p^n$$
2. ...immer scheitert?
$$(1-p) \cdot (1-p) \cdots (1-p) = (1-p)^n$$
3. ...erst k mal gelingt und dann $n-k$ mal scheitert?
$$p^k \cdot (1-p)^{n-k}$$
4. ...insgesamt k mal gelingt und $n-k$ mal scheitert?
$$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

Erläuterung

$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$ ist die Anzahl der Möglichkeiten, die k Erfolge in die n Versuche einzusortieren.

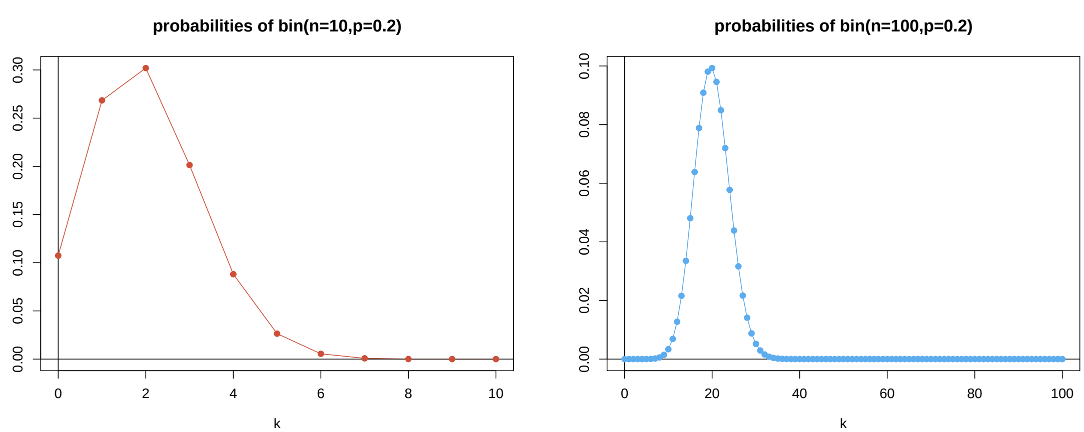
Binomialverteilung

Sei X die Anzahl der Erfolge bei n unabhängigen Versuchen mit Erfolgswahrscheinlichkeit von jeweils p. Dann gilt für $k \in \{0, 1, \ldots, n\}$

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

und X heißt *binomialverteilt*, kurz:

$$X \sim \text{bin}(n, p).$$



Mit der Binomialverteilung können wir nun eine unserer anfänglichen Fragen bearbeiten:

Nehmen wir an, in einer kleinen Population von $n = 100$ Individuen hat ein neutrales Allel A in der aktuellen Generation eine Häufigkeit von 0.32.

Wie wahrscheinlich ist es, dass die Häufigkeit X von A in der nächsten Generation wieder genau gleich ist?

$$\Pr(X = 0.32) = ?$$

Dazu müssen wir die Population zunächst stochastisch modellieren, und das Ergebnis wird davon abhängen, wie wir das machen.

Modellierungsansatz

Wir gehen von folgenden vereinfachenden Annahmen aus:

- Diskrete Generationen.
- Die Population ist haploid, d.h. jedes Individuum hat genau ein Elter in der Generation davor.
- Konstante Populationsgröße von $n = 100$

Auch dann hängt $\Pr(X = 0.32)$ noch davon ab, ob wenige Individuen ganz viele Nachkommen haben oder alle ungefähr gleich viele. Um $\Pr(X = 0.32)$ brauchen wir eine weitere Modellannahme, z.B.:

- Jedes Individuum wählt rein zufällig einen Elter in der vorherigen Generation.

"rein zufällig" bedeutet *unabhängig von allen anderen* und *alle potentiellen Eltern mit der selben Wahrscheinlichkeit*.

Aus unseren Annahmen folgt, dass alle Individuen der nächsten Generation unabhängig von einander mit der selben Wahrscheinlichkeit von jeweils 0.32 das Allel A haben.

Also ist die Anzahl K der Individuen der nächsten Generation mit Allel A binomial verteilt mit $n = 100$ und $p = 0.32$:

$$K \sim \mathrm{bin}(n = 100, p = 0.32)$$

Es folgt für $X = K/n$:

$$\Pr(X = 0.32) = \Pr(K = 32) = \binom{n}{32} \cdot p^{32} \cdot (1-p)^{100-32}$$

$$= \binom{100}{32} \cdot 0.32^{32} \cdot 0.68^{68} \approx 0.085$$

Bildquellen

`http://de.wikipedia.org/w/index.php?title=Datei:Falling_ball.jpg&filetimestamp=20071020133134`
[(c) by Michael Maggs]

Was Sie u.a. erklären können sollten

- Konzepte und Schreibweisen: Ereignisse, Zufallsvariablen und Wahrscheinlichkeitsverteilungen
- Einschluss-Ausschluss-Formel
- Wie man eine Funktion auf eine Zufallsvariable anwendet
- Bedingte Wahrscheinlichkeiten
- Stochastische Unabhängigkeit von Ereignissen bzw. Zufallsvariablen
- Binomialverteilung und $\binom{n}{k}$

4 Erwartungswert

Beispiel: In einer Klasse mit 30 Schülern gab es in der letzten Klausur 3x Note 1, 6x Note 2, 8x Note 3, 7x Note 4, 4x Note 5 und 2x Note 6.
Die Durchschnittsnote war

$$\frac{1}{30}\Big(1{+}1{+}1 \;+\; \underbrace{2{+}\ldots{+}2}_{6\text{ mal}} \;+\; \underbrace{3{+}\ldots{+}3}_{8\text{ mal}}$$
$$+\; \underbrace{4{+}\ldots{+}4}_{7\text{ mal}} \;+\; 5{+}\,5{+}\,5{+}5 \;+\; 6{+}6\Big) = \frac{1}{30}99 = 3.3$$

Mit relativen Häufigkeiten:

$$1\cdot\frac{3}{30}+2\cdot\frac{6}{30}+3\cdot\frac{8}{30}+4\cdot\frac{7}{30}+5\cdot\frac{4}{30}+6\cdot\frac{2}{30}=3.3$$

Merke: Der Durchschnittswert ist die Summe über alle möglichen Werte gewichtet mit den relativen Häufigkeiten
Analog zum Durchschnittswert gibt es bei Zufallsvariablen der Erwartungswert.

Definition 2 (Erwartungswert) *Sei X eine Zufallsvariable mit endlichem oder abzählbarem Wertebereich $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$. Dann ist der Erwartungswert von X definiert durch*

$$\mathbb{E}X = \sum_{x\,\in\,\mathcal{S}} x\cdot\Pr(X=x)$$

Man schreibt auch μ_X statt $\mathbb{E}X$.
Beispiele:

- Wenn $\mathcal{S}$ eine endliche Menge ist und alle x gleich wahrscheinlich sind, so erhält man die bekannte Formel

$$\text{Erwartungswert} = \frac{\text{Summe der Werte}}{\text{Anzahl der Werte}}.$$

- Ist k_x die Häufigkeit des Wertes x in einer Gesamtheit der Größe n, aus der rein zufällig gezogen wird, so folgt:

$$\mathbb{E}X = \sum_x x\cdot\frac{k_x}{n} = \frac{\sum_x x\cdot k_x}{n}$$

Definition 3 (Erwartungswert) *Sei X eine Zufallsvariable mit endlichem oder abzählbarem Wertebereich $\mathcal{S} = \{x_1, x_2, x_3 \ldots\} \subseteq \mathbb{R}$. Dann ist der Erwartungswert von X definiert durch*

$$\mathbb{E}X = \sum_{x\,\in\,\mathcal{S}} x\cdot\Pr(X=x)$$

Beispiele:

- Sei X eine 0-1-Zufallsvariable mit Erfolgswahrscheinlichkeit $p\in[0,1]$. Dann gilt

$$\mathbb{E}X = 1\cdot\Pr(X=1)+0\cdot\Pr(X=0)=\Pr(X=1)=p$$

- Sei W die Augenzahl bei einem Würfelwurf. Dann gilt

$$\begin{aligned}\mathbb{E}W &= 1\cdot\Pr(W=1)+2\cdot\Pr(W=2)+\ldots+6\cdot\Pr(W=6)\\ &= 1\cdot\tfrac{1}{6}+2\cdot\tfrac{1}{6}+\ldots+6\cdot\tfrac{1}{6} = 21\tfrac{1}{6} = 3.5\end{aligned}$$

Satz 1 (Erwartungswert von $f(X)$) *Sei X eine Zufallsvariable mit endlichem oder abzählbarem Wertebereich $\mathcal{S}$. Sei $f\colon \mathcal{S}\to\mathbb{R}$ eine Funktion. Dann ist der Erwartungswert von $f(X)$ durch $\mathbb{E}[f(X)] = \sum_{y\,\in\,f(\mathcal{S})} y\cdot\Pr(f(X)=y)$ definiert und es gilt*

$$\mathbb{E}[f(X)] = \sum_{x\,\in\,\mathcal{S}} f(x)\cdot\Pr(X=x).$$

Beispiel: Auf die Augenzahl W bei einem Würfelwurf wenden wir die Funktion $f(x)=(x-4)^2$ an. Dann hat $f(W)$ die möglichen Werte 0, 1, 4 und 9, und nach Definition ist:

$$\mathbb{E}[(W-4)^2] = \sum_y y\cdot\Pr(f(W)=y) = 0\cdot\frac{1}{6}+1\cdot\frac{1}{3}+4\cdot\frac{1}{3}+9\cdot\frac{1}{6}$$

Nach obigem Satz kann man aber auch so rechnen:

$$\mathbb{E}[(W-4)^2] = (1-4)^2\cdot\tfrac{1}{6}+(2-4)^2\cdot\tfrac{1}{6}+\ldots+(6-4)^2\cdot\tfrac{1}{6}$$

Wieso Beweise?

- Besseres Verständnis für das Konzept des Erwartungswerts
- Verständnis, wieso die Regeln gelten und andere nicht
- Regeln nicht auswendig lernen sondern verstehen/herleiten können
- Mal wieder Mathe üben
- Als Wissenschaftler dürfen Sie nichts ohne Beweis/Evidenz glauben!

Beweis von $\mathbb{E}[f(X)] = \sum_{x \in \mathcal{S}} f(x) \cdot \Pr(X = x)$:

Sei $f(S)$ das **Bild** von f, also die Menge aller $f(x)$ mit $x \in S$.

$$\begin{aligned}
\mathbb{E}[f(X)] &= \sum_{y \in f(\mathcal{S})} y \cdot \Pr(f(X) = y) \\
&= \sum_{y \in f(\mathcal{S})} y \cdot \Pr(X \in f^{-1}(\{y\})) \\
&= \sum_{y \in f(\mathcal{S})} \sum_{x \in f^{-1}(\{y\})} y \cdot \Pr(X = x) \\
&= \sum_{y \in f(\mathcal{S})} \sum_{x \in f^{-1}(\{y\})} f(x) \cdot \Pr(X = x) \\
&= \sum_{x \in S} f(x) \cdot \Pr(X = x)
\end{aligned}$$

(Achtung: f muss nicht umkehrbar sein, denn $f^{-1}(\{y\})$ ist das Urbild der Menge $\{y\}$ und damit selbst eine Menge.)

Rechnen mit Erwartungswerten

Satz 2 (Linearität des Erwartungswerts) *Sind X und Y Zufallsvariablen mit Werten in $\mathbb{R}$ und ist $a \in \mathbb{R}$, so gilt:*

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$

Satz 3 (Nur für Unabhängige!) *Sind X und Y **stochastisch unabhängige** Zufallsvariablen mit Werten in $\mathbb{R}$, so gilt*

- $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y$.

Im allgemeinen gilt $\mathbb{E}(X \cdot Y) \neq \mathbb{E}X \cdot \mathbb{E}Y$. Beispiel:

$$\mathbb{E}(W \cdot W) = \tfrac{91}{6} = 15.167 > 12.25 = 3.5 \cdot 3.5 = \mathbb{E}W \cdot \mathbb{E}W$$

Beweis von $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$:

Mit $f(x) = a \cdot x$ erhalten wir aus dem zuvor bewiesenen Satz:

$$\begin{aligned}
\mathbb{E}(a \cdot X) &= \sum_x f(x) \cdot \Pr(X = x) \\
&= \sum_x a \cdot x \cdot \Pr(X = x) \\
&= a \cdot \sum_x x \cdot \Pr(X = x) \\
&= a \cdot \mathbb{E}X
\end{aligned}$$

$\square$

Wenn Y eine Zufallsvariable mit diskretem Zustandsraum $\mathcal{S}$ ist, und X eine weitere Zufallsvariable, so gilt

$$\sum_{y\in\mathcal{S}} \Pr(X=x, Y=y) = \Pr(X=x, Y\in\mathcal{S}) = \Pr(X=x).$$

Das werden wir im nachfolgenden Beweises verwenden.
Beweis von $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$: Wir gehen zur vereinfachend davon aus, dass X und Y den selben Zustandsraum $\mathcal{S}$ haben.

Auch hier wenden wir den Satz wie im vorherigen Beweis an, diesmal mit $f(\ (x,y)\ ) = x+y$. Damit folgt:

$$\begin{aligned}
\mathbb{E}(X+Y) = \mathbb{E}(f(X,Y)) &= \sum_{(x,y)\in\mathcal{S}^2} f(x,y)\cdot\Pr((X,Y)=(x,y)) \\
&= \sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} (x+y)\cdot\Pr(X=x,Y=y) \\
&= \sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} x\cdot\Pr(X=x,Y=y) + \sum_{y\in\mathcal{S}}\sum_{x\in\mathcal{S}} y\cdot\Pr(X=x,Y=y) \\
&= \sum_{x\in\mathcal{S}} x\cdot\sum_{y\in\mathcal{S}}\Pr(X=x,Y=y) + \sum_{y\in\mathcal{S}} y\cdot\sum_{x\in\mathcal{S}}\Pr(X=x,Y=y) \\
&= \sum_{x\in\mathcal{S}} x\cdot\Pr(X=x) + \sum_{y\in\mathcal{S}} y\cdot\Pr(Y=y) \\
&= \mathbb{E}(X)+\mathbb{E}(Y)
\end{aligned}$$

$\square$

Beweis der Produktformel: Sei $\mathcal{S}$ der Zustandsraum von X und Y und seien X und Y (stochastisch) unabhängig.

$$\begin{aligned}
&\mathbb{E}(X\cdot Y) \\
&= \sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} (x\cdot y)\Pr(X=x,Y=y) \\
&= \sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}} (x\cdot y)\Pr(X=x)\Pr(Y=y) \\
&= \sum_{x\in\mathcal{S}} x\Pr(X=x)\cdot\sum_{y\in\mathcal{S}} y\Pr(Y=y) \\
&= \mathbb{E}X\cdot\mathbb{E}Y\cdot
\end{aligned}$$

$\square$

Erwartungswert der Binomialverteilung

Seien $Y_1, Y_2, \ldots, Y_n$ die Indikatorvariablen der n unabhängigen Versuche d.h.

$$Y_i = \begin{cases} 1 & \text{falls der } i-\text{te Versuch gelingt} \\ 0 & \text{falls der } i-\text{te Versuch scheitert} \end{cases}$$

Dann ist $X = Y_1+\cdots+Y_n$ binomialverteilt mit Parametern (n,p), wobei p die Erfolgswahrscheinlichkeit der Versuche ist.

Wegen der Linearität des Erwartungswerts gilt

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}(Y_1+\cdots+Y_n) \\
&= \mathbb{E}Y_1+\cdots+\mathbb{E}Y_n \\
&= p+\cdots+p \ = \ np
\end{aligned}$$

Wir halten fest:

$$X\sim\text{bin}(n,p)\Rightarrow\mathbb{E}X = n\cdot p$$

5 Varianz und Korrelation

Definition 4 (Varianz, Kovarianz und Korrelation) *Die Varianz einer $\mathbb{R}$-wertigen Zufallsgröße X ist*

$$\mathit{Var}X = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}X)^2\right].$$

$\sigma_X = \sqrt{\mathit{Var}\ X}$ *ist die Standardabweichung.*
Ist Y eine weitere reellwertige Zufallsvariable, so ist

$$\mathit{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}X)\cdot(Y - \mathbb{E}Y)\right]$$

die Kovarianz von X und Y.
Die Korrelation von X und Y ist

$$\mathit{Cor}(X,Y) = \frac{\mathit{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}.$$

Die Varianz

$$\mathrm{Var}X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

ist die durchschnittliche quadrierte Abweichung vom Mittelwert.

Die Korrelation

$$\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\cdot\sigma_Y}$$

liegt immer im Intervall $[-1,1]$. Die Variablen X und Y sind

- positiv korreliert, wenn X und Y tendenziell entweder beide überdurchschnittlich große Werte oder beide unterdurchschnittlich große Werte annehmen.
- negativ korreliert, wenn X und Y tendenziell auf verschiedenen Seiten ihrer Erwartungswerte liegen.

Sind X und Y unabhängig, so sind sie auch unkorreliert, d.h. $\mathrm{Cor}(X,Y) = 0$.

Beispiel: Würfel

Varianz des Würfelergebnisses W:

$$\begin{aligned}
\mathrm{Var}(W) &= \mathbb{E}\big[\big(W - \mathbb{E}W\big)^2\big] \\
&= \mathbb{E}\big[\big(W - 3.5\big)^2\big] \\
&= (1-3.5)^2\cdot\frac{1}{6} + (2-3.5)^2\cdot\frac{1}{6} + \ldots + (6-3.5)^2\cdot\frac{1}{6} \\
&= \frac{17.5}{6} \\
&= 2.91667
\end{aligned}$$

Beispiel: Die empirische Verteilung

Sind $x_1,\ldots,x_n \in \mathbb{R}$ Daten und entsteht X durch rein zufälliges Ziehen aus diesen Daten (d.h. $\Pr(X = x) = \frac{n_x}{n}$, wobei $n_x = |\{i\ :\ x_i = x\}|$ die Anzahl der x_i ist mit $x_i = x$), so gilt:

$$\mathbb{E}X = \sum_x x\cdot\frac{n_x}{n} = \sum_{i=1}^n x_i\cdot\frac{1}{n} = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}$$

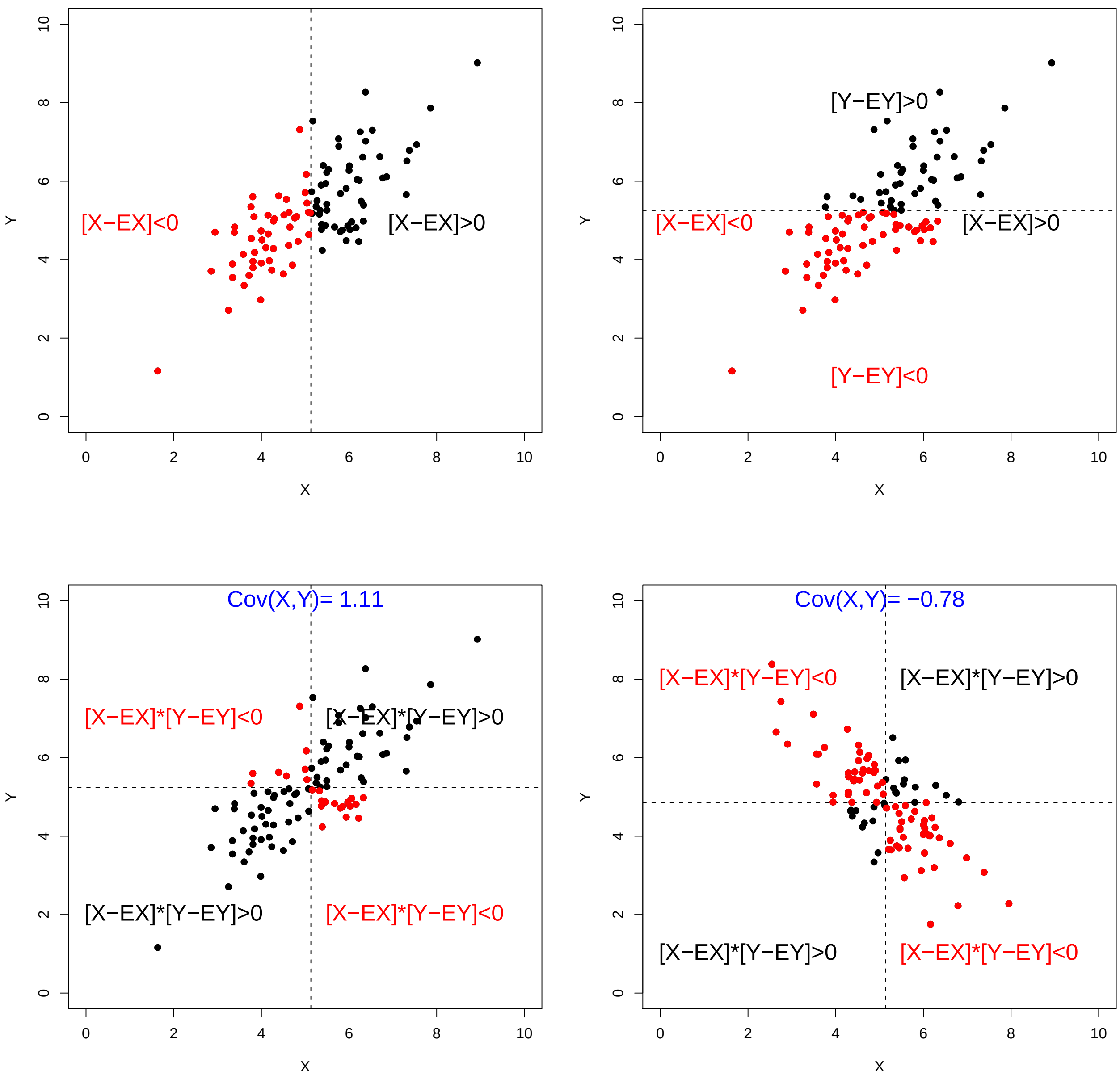
und

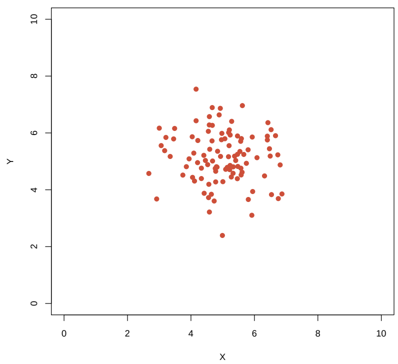
$$\mathrm{Var}\ X = \mathbb{E}\big[\big(X - \mathbb{E}X\big)^2\big] = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$$

Sind $(x_1,y_1),\ldots,(x_n,y_n) \in \mathbb{R}\times\mathbb{R}$ Daten und entsteht (X,Y) durch rein zufälliges Ziehen aus diesen Daten (d.h. $\Pr((X,Y) = (x,y)) = \frac{|\{i\ :\ (x_i,y_i)=(x,y)\}|}{n}$), so gilt:

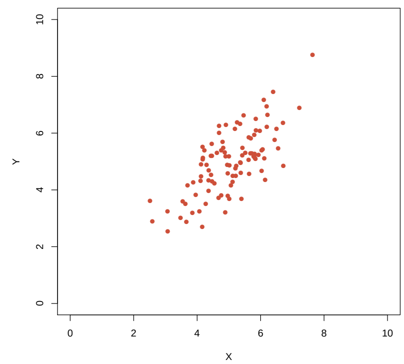
$$\mathrm{Cov}\ (X,Y) = \mathbb{E}\big[\big(X - \mathbb{E}X\big)\big(Y - \mathbb{E}Y\big)\big] = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})$$

Wieso Cov$(X, Y) = \mathbb{E}([X - \mathbb{E}X][Y - \mathbb{E}Y])$**?**

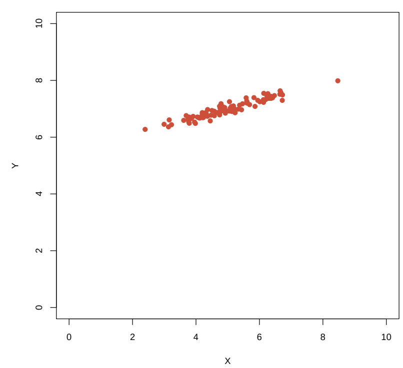




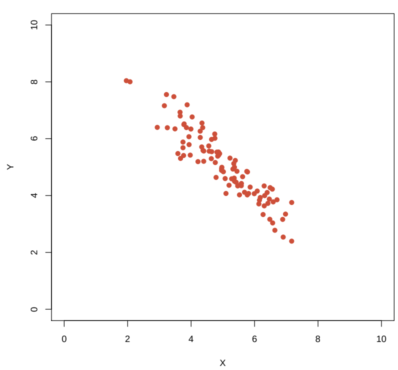
$\sigma_X = 0.95$, $\sigma_Y = 0.92$
$\mathrm{Cov}(X, Y) = -0.06$
$\mathrm{Cor}(X, Y) = -0.069$



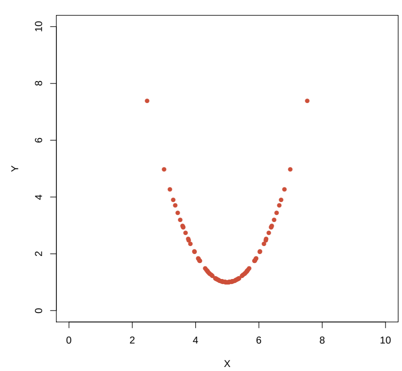
$\sigma_X = 1.14$, $\sigma_Y = 0.97$
$\mathrm{Cov}(X, Y) = 0.78$
$\mathrm{Cor}(X, Y) = 0.71$



$\sigma_X = 1.03$, $\sigma_Y = 0.32$
$\mathrm{Cov}(X, Y) = 0.32$
$\mathrm{Cor}(X, Y) = 0.95$

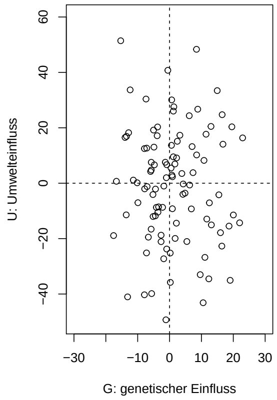


$\sigma_X = 1.13$, $\sigma_Y = 1.2$
$\mathrm{Cov}(X, Y) = -1.26$
$\mathrm{Cor}(X, Y) = -0.92$



$\sigma_X = 0.91$, $\sigma_Y = 0.88$
$\mathrm{Cov}(X, Y) = 0$
$\mathrm{Cor}(X, Y) = 0$

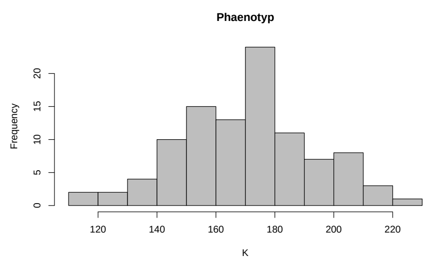
Wieso var und cov und nicht nur σ und cor? Ein Beispiel




```
> sd(G)
[1] 9.49879
> sd(U)
[1] 20.77562
```

Phänotyp:

```
K <- 170 + G + U
```



Welcher Anteil von K ist genetisch bedingt?

```
> sd(G)/sd(K)
[1] 0.4230486

> sd(U)/sd(K)
[1] 0.9252861

> var(G)/var(K)
[1] 0.1789701
> var(U)/var(K)
[1] 0.8561545

> var(K)
[1] 504.1457
> var(G) + var(U)
[1] 521.8536
> cov(G,U)
[1] -8.853946
> var(G) + var(U) + 2*cov(G,U)
[1] 504.1457
```

Rechenregeln für Kovarianzen

$\text{Cov}(X,Y) = \mathbb{E}[(X - EX) \cdot (Y - \mathbb{E}Y)]$

- Sind X und Y unabhängig, so folgt $\text{Cov}(X,Y) = 0$ (die Umkehrung gilt nicht!)
- $\text{Cov}(X,Y)$=$\text{Cov}(Y,X)$
- $\text{Cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$
- $\text{Cov}(a \cdot X, Y) = a \cdot \text{Cov}(X,Y) = \text{Cov}(X, a \cdot Y)$
- $\text{Cov}(X + Z, Y) = \text{Cov}(X,Y) + \text{Cov}(Z,Y)$
- $\text{Cov}(X, Z + Y) = \text{Cov}(X,Z) + \text{Cov}(X,Y)$

Die letzten drei Regeln beschreiben die Bilinearität der Kovarianz.

Beweise an der Tafel; Wieso schon wieder Beweise? selbe Gründe wie vorhin + Anwendungsbeispiele für Regeln für $\mathbb{E}$

Rechenregeln für Varianzen

$\text{Var}X = \mathbb{E}[(X - \mathbb{E}X)^2]$

- $\text{Var}X = \text{Cov}(X,X)$
- $\text{Var}X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
- $\text{Var}(a \cdot X) = a^2 \cdot \text{Var}X$
- $\text{Var}(X + Y) = \text{Var}X + \text{Var}Y + 2 \cdot \text{Cov}(X,Y)$
- $\text{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \text{Var}\big(X_i\big) + 2 \cdot \sum_{j=1}^n \sum_{i=1}^{j-1} \text{Cov}(X_i, X_j)$
- Sind (X,Y) stochastisch unabhängig, so folgt:

$$\text{Var}(X + Y) = \text{Var}X + \text{Var}Y$$

(Beweise an der Tafel; ergeben sich aus den Rechenregeln für Kovarianzen)

Rechenregeln für die Korrelation

$\text{Cor}(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$

- $-1 \leq \text{Cor}(X,Y) \leq 1$
- $\text{Cor}(X,Y) = \text{Cor}(Y,X)$
- $\text{Cor}(X,Y) = \text{Cov}(X/\sigma_X, Y/\sigma_Y)$
- $\text{Cor}(X,Y) = 1$ genau dann wenn Y eine wachsende, affin-lineare Funktion von X ist, d.h. falls es $a > 0$ und $b \in \mathbb{R}$ gibt, so dass $Y = a \cdot X + b$
- $\text{Cor}(X,Y) = -1$ genau dann wenn Y eine fallende, affin-lineare Funktion von X ist, d.h. falls es $a < 0$ und $b \in \mathbb{R}$ gibt, so dass $Y = a \cdot X + b$

Bernoulli-Verteilung

Eine Bernoulli-verteilte Zufallsvariable Y mit Erfolgsws $p \in [0,1]$ hat Erwartungswert

$$\mathbb{E}Y = p$$

und Varianz

$$\text{Var } Y = p \cdot (1-p)$$

Beweis: Aus $\Pr(Y=1) = p$ und $\Pr(Y=0) = (1-p)$ folgt

$$\mathbb{E}Y = 1 \cdot p + 0 \cdot (1-p) = p.$$

Varianz:

$$\begin{aligned}\text{Var } Y &= \mathbb{E}\big(Y^2\big) - \big(\mathbb{E}Y\big)^2 \\ &= 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p \cdot (1-p)\end{aligned}$$

Binomialverteilung

Seien nun $Y_1, \cdots, Y_n$ unabhängig und Bernoulli-verteilt mit Erfolgsws p. Dann gilt

$$\sum_{i=1}^{n} Y_i =: X \sim \text{bin}(n,p)$$

und es folgt:

$$\text{Var } X = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{Var } Y_i = n \cdot p \cdot (1-p)$$

Binomialverteilung

Satz 4 (Erwartungswert und Varianz der Binomialverteilung) *Ist X binomialverteilt mit Parametern (n,p), so gilt:*

$$\mathbb{E}X = n \cdot p$$

und

$$Var\ X = n \cdot p \cdot (1-p)$$

Beispiel: Genetische Drift

In einer haploiden Population von n Individuen habe Allel A die Häufigkeit p. Wir gehen wieder davon aus, dass jedes der n Individuen der nächsten Generation unabhängig von den anderen mit W'keit p Allel A bekommt. Die absolute Häufigkeit K von A in der nächsten Generation ist damit (n, p)-binomialverteilt.

Dann folgt für die relative Häufigkeit $X = K/n$ in der nächsten Generation:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(K/n) = \mathrm{Var}(K)/n^2 = n \cdot p \cdot (1-p)/n^2 \\ &= \frac{p \cdot (1-p)}{n}\end{aligned}$$

Beispiel: Genetische Drift

Betrachtet man die Änderung der Allelhäufigkeit über m Generationen, so addieren sich die Varianzen. Ist m eine kleine Zahl, so dass es nur um so wenige Generationen geht, dass sich p nur geringfügig ändert, hat die Änderung der Genhäufigkeit ungefähr die Varianz

$$m \cdot \mathrm{Var}(X) = \frac{m \cdot p \cdot (1-p)}{n}$$

(weil die Zuwächse der einzelnen Generationen von einander unabhängig sind) und damit die Standardabweichung

$$\sqrt{m} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Mit den Rechenregeln können wir nun auch endlich beweisen:

Satz 5 *Sind $X_1, X_2, \ldots X_n$ unabhängige $\mathbb{R}$-wertige Zufallsgrößen mit Mittelwert μ und Varianz σ^2, so gilt für $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:*

$$\mathbb{E}\overline{X} = \mu$$

und

$$Var\ \overline{X} = \frac{1}{n}\sigma^2,$$

d.h.

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Insbesondere: Der Standardfehler $\frac{s}{\sqrt{n}}$ ist ein Schätzer der Standardabweichung $\sigma_{\overline{X}}$ des Stichprobenmittels $\overline{X}$ der Stichprobe $(X_1, X_2, \ldots, X_n)$.
Die Stichproben-Standardabweichung s ist ein Schätzer der Standardabweichung σ der Grundgesamtheit.

Beweis: Linearität des Erwartungswertes impliziert

$$\begin{aligned}\mathbb{E}\overline{X} &= \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big(X_i\big) \\ &= \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.\end{aligned}$$

Die Unabhängigkeit der X_i vereinfacht die Varianz zu

$$\begin{aligned}\mathrm{Var}\ \overline{X} &= \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^{n} X_i\Big) = \frac{1}{n^2}\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\big(X_i\big) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}\sigma^2\end{aligned}$$

Wir können nun außerdem beweisen, dass die korrigierte Stichprobenvarianz als Schätzer für die Varianz erwartungstreu (engl. *unbiased*) ist:

Satz 6 *Sind $X_1, X_2, \dots X_n$ unabhängige $\mathbb{R}$-wertige Zufallsgrößen mit Mittelwert μ und Varianz σ^2, so gilt für $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$:*

$$\mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2\right) = \sigma^2$$

Beweis: Zunächst machen wir klar, dass für $i \neq j$ gilt:

$$\mathbb{E}X_i = \mathbb{E}X_j = \mathbb{E}X_1 \qquad \text{und} \qquad \mathbb{E}X_i \cdot X_j = \mathbb{E}X_i \cdot \mathbb{E}X_j = \left(\mathbb{E}X_1\right)^2$$

also gilt

$$\begin{aligned}
\mathbb{E}\left(X_i \cdot \sum_{j=1}^n X_j\right) &= \mathbb{E}\left(X_i \cdot \left(X_i + \sum_{j\neq i} X_j\right)\right) = \mathbb{E}\left(X_i \cdot X_i\right) + \mathbb{E}\sum_{j\neq i} X_i \cdot X_j \\
&= \mathbb{E}\left(X_1^2\right) + \sum_{j\neq i}\mathbb{E}\left(X_i \cdot X_j\right) = \mathbb{E}\left(X_1^2\right) + \sum_{j\neq i}\mathbb{E}X_1 \cdot \mathbb{E}X_2 \\
&= \mathbb{E}\left(X_1^2\right) + (n-1)\cdot\mathbb{E}X_1 \cdot \mathbb{E}X_2 \\
&= \mathbb{E}\left(X_1^2\right) + (n-1)\cdot\left(\mathbb{E}X_1\right)^2
\end{aligned}$$

Mit $\mathbb{E}\left(X_i \cdot \sum_{j=1}^n X_j\right) = \mathbb{E}X_1^2 + (n-1)\cdot\left(\mathbb{E}X_1\right)^2$ folgt nun (Achtung: $\mathbb{E}X^2$ heißt $\mathbb{E}(X^2)$, nicht $(\mathbb{E}X)^2$):

$$\begin{aligned}
\mathbb{E}\left(X_i - \overline{X}\right)^2 &= \mathbb{E}\left(X_i - \frac{1}{n}\sum_j X_j\right)^2 \\
&= \mathbb{E}\left(X_i^2 - \frac{2}{n}X_i \cdot \sum_j X_j + \frac{1}{n^2}\cdot\sum_j X_j \cdot \sum_k X_k\right) \\
&= \mathbb{E}X_i^2 - \frac{2}{n}\mathbb{E}\left(X_i \cdot \sum_j X_j\right) + \frac{1}{n^2}\sum_j \mathbb{E}\left(X_j \cdot \sum_k X_k\right) \\
&= \mathbb{E}X_1^2 + \left(-\frac{2}{n} + \frac{n}{n^2}\right)\cdot\left(\mathbb{E}\left(X_1^2\right) + (n-1)\cdot\left(\mathbb{E}X_1\right)^2\right) \\
&= \frac{n-1}{n}\left(\mathbb{E}X_1^2 - \left(\mathbb{E}X_1\right)^2\right) = \frac{n-1}{n}\mathrm{Var}X_1 = \frac{n-1}{n}\sigma^2
\end{aligned}$$

Mit $\mathbb{E}\left(X_i - \overline{X}\right)^2 = \frac{n-1}{n}\sigma^2$ erhalten wir:

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2\right) &= \frac{1}{n-1}\sum_{i=1}^n \mathbb{E}\left(\left(X_i - \overline{X}\right)^2\right) \\
&= \frac{n}{n-1}\mathbb{E}\left(\left(X_i - \overline{X}\right)^2\right) \\
&= \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2
\end{aligned}$$

□

Aber Achtung!

Aus der Erwartungstreue des korrgierten Schätzers für die Stichprobenvarianz folgt nicht, dass die entsprechende Schätzung für die Standardabweichung Erwartungstreu wäre. Ist $\sigma > 0$, so gilt:

$$\mathbb{E}\sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2} < \sqrt{\mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2\right)} = \sigma$$

Was Sie u.a. erklären können sollten

- Definitionen von $\mathbb{E}$, Var, Cov, Cor für Zufallsvariablen
- Rechenregeln für $\mathbb{E}$, Var, Cov, Cor und wie man sie benutzt
- Unterschied zwischen Korrelation und stochastischer Abhängigkeit
- $\mathbb{E}$ und Var (und SD) der Binomialverteilung
- Wie genetische Drift von der Populationsgröße und der Allelhäufigkeit abhängt und wie sich das aus den Eigenschaften der Binomialverteilung ergibt
- Grundlegende Prinzipien der Beweise in diesem Abschnitt

6 Ein Anwendungsbeispiel

In

Literatur

[1] E.N. Moriyama (2003) Codon Usage *Encyclopedia of the human genome*, Macmillan Publishers Ltd.

werden u.a. 9497 menschliche Gene auf "Codon Bias" untersucht.
In diesen Genen wird die Aminosäure Prolin 16710 mal durch das Codon CCT und 18895 mal durch das Codon CCC codiert.
Ist es nur vom reinen Zufall abhängig, welches Codon verwendet wird?
Dann wäre die Anzahl X der CCC binomialverteilt mit $p = \frac{1}{2}$ und $n = 16710 + 18895 = 35605$. Angenommen die Anzahl X $(= 18895)$ der CCC ist binomialverteilt mit $p = \frac{1}{2}$ und $n = 16710 + 18895 = 35605$.

$$\mathbb{E}X = n \cdot p = 17802.5$$
$$\sigma_X = \sqrt{n \cdot p \cdot (1-p)} \approx 94.34$$
$$18895 - 17802.5 = 1092.5 \approx 11.6 \cdot \sigma_X$$

Sieht das nach Zufall aus?

Die Frage ist:

Wie groß ist die Wahrscheinlichkeit einer Abweichung vom Erwartungswert von mindestens $\approx 11.6 \cdot \sigma_X$, wenn alles Zufall ist?

Wir müssen also

$$\Pr\big(|X - \mathbb{E}X| \geq 11.6\sigma_X\big)$$

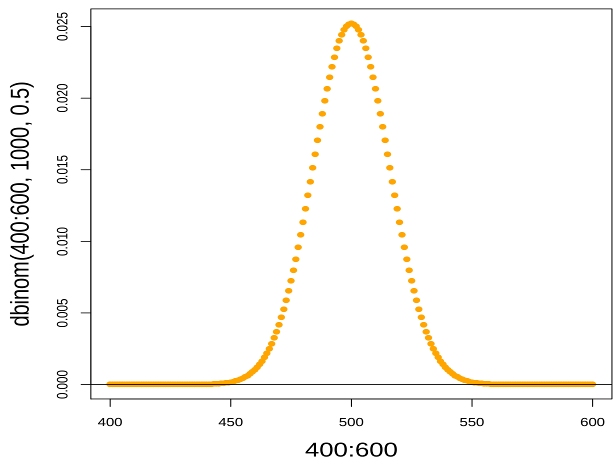
berechnen.

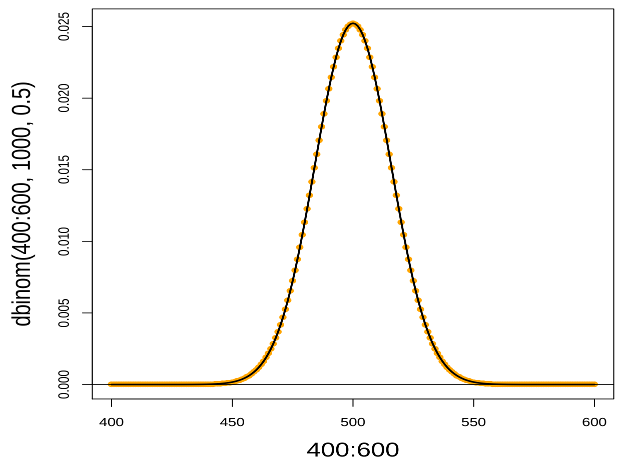
Das Problem bei der Binomialverteilung ist: $\binom{n}{k}$ exakt zu berechnen, ist für große n sehr aufwändig. Deshalb:

Die Binomialverteilung wird oft durch andere Verteilungen approximiert.

7 Die Normalverteilung

Die Binomialverteilung mit großer Versuchzahl n sieht aus wie die Normalverteilung:



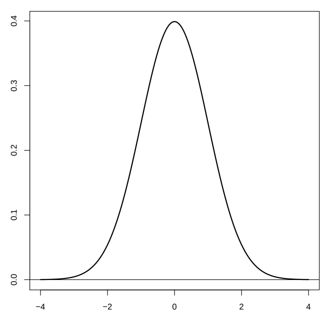


Dichte der Standardnormalverteilung

Eine Zufallsvariable Z mit der Dichte

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

heißt *standardnormalverteilt*.



kurz: $Z \sim \mathcal{N}(0,1)$

$$\mathbb{E}Z = 0$$

$$\text{Var } Z = 1$$

"Gauß-Glocke"

Ist Z $\mathcal{N}(0,1)$-verteilt, so ist $X = \sigma \cdot Z + \mu$ normalverteilt mit Mittelwert μ und Varianz σ^2, kurz:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

X hat dann die Dichte

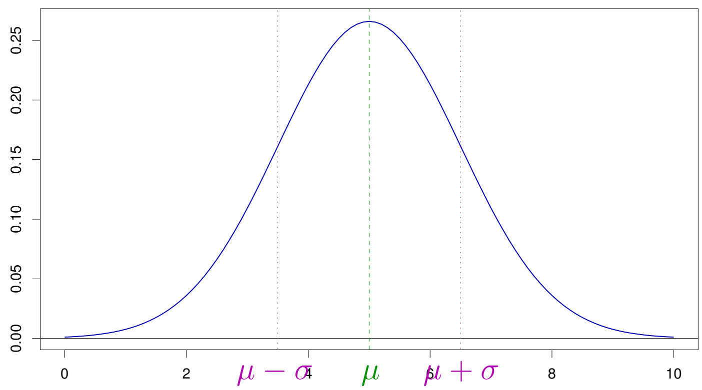
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Merkregeln

Ist $Z \sim \mathcal{N}(\mu, \sigma^2)$, so gilt:

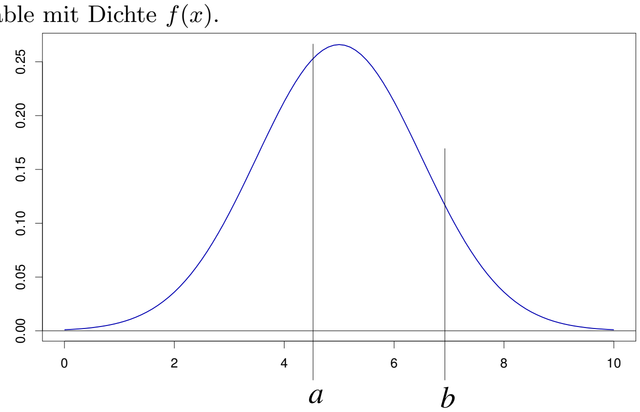
- $\Pr(|Z - \mu| > \sigma) \approx 33\%$
- $\Pr(|Z - \mu| > 1.96 \cdot \sigma) \approx 5\%$
- $\Pr(|Z - \mu| > 3 \cdot \sigma) \approx 0.3\%$

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



Dichten brauchen Integrale

Sei Z eine Zufallsvariable mit Dichte $f(x)$.



Dann gilt

$$\Pr(Z \in [a, b]) = \int_a^b f(x) \mathsf{d}x.$$

Frage: Wie berechnet man $\Pr(Z = 5)$?

Antwort: Für jedes $x \in \mathbb{R}$ gilt $\Pr(Z = x) = 0$ (da Fläche der Breite 0)

Was wird dann aus $\mathbb{E}Z = \sum_{x \in \mathcal{S}} x \cdot \Pr(Z = x)$?

Bei einer kontinuierlichen Zufallsvariable mit Dichtefunktion f definiert man:

$$\mathbb{E}Z := \int_{-\infty}^{\infty} x \cdot f(x) \mathsf{d}x$$

Die Normalverteilung in *R*

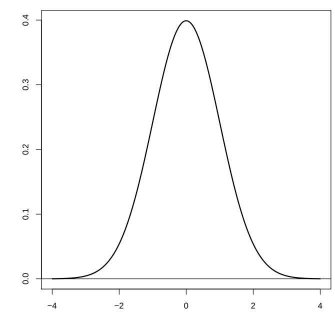
Die Normalverteilung hat in R das Kürzel *'norm'*.

Es gibt 4 R-Befehle:

dnorm():	Dichte der Normalverteilung (**d**ensity)
rnorm():	Ziehen einer Stichprobe (**r**andom sample)
pnorm():	Verteilungsfunktion der Normalverteilung (**p**robability)
qnorm():	Quantilfunktion der Normalverteilung (**q**uantile)

Beispiel: Dichte der Standardnormalverteilung:

`> plot(dnorm,from=-4,to=4)`



`> dnorm(0) [1] 0.3989423 > dnorm(0,mean=1,sd=2) [1] 0.1760327`

Beispiel: Ziehen einer Stichprobe

Ziehen einer Stichprobe der Länge 6 aus einer Standardnormalverteilung:

`> rnorm(6) [1] -1.24777899 0.03288728 0.19222813 0.81642692 -0.62607324 -1.09273888`

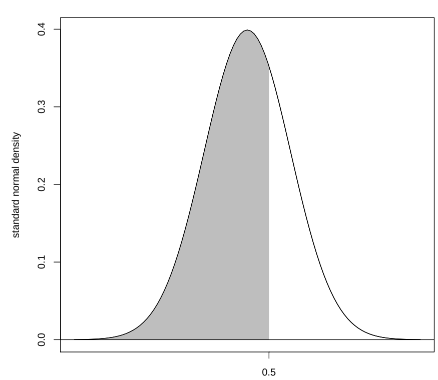
Ziehen einer Stichprobe der Länge 7 aus einer Normalverteilung mit Mittelwert 5 und Standardabweichung 3:

`> rnorm(7,mean=5,sd=3) [1] 2.7618897 6.3224503 10.8453280 -0.9829688 5.6143127 0.6431437 8.123570`

Beispiel: Berechnung von Wahrscheinlichkeiten: Sei $Z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$, also standardnormalverteilt.

$\Pr(Z < a)$ berechnet man in R mit `pnorm(a)`

`> pnorm(0.5) [1] 0.6914625`

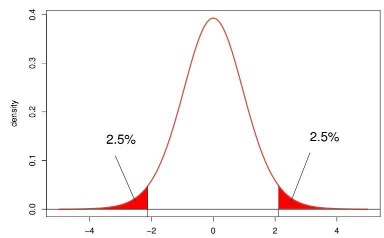


Beispiel: Berechnung von Wahrscheinlichkeiten: Sei $Z \sim \mathcal{N}(\mu = 5, \sigma^2 = 2.25)$.

Berechnung von $\Pr(Z \in [3, 4])$:

$$\Pr(Z \in [3, 4]) = \Pr(Z < 4) - \Pr(Z < 3)$$

`> pnorm(4,mean=5,sd=1.5)-pnorm(3,mean=5,sd=1.5) [1] 0.1612813` **Beispiel**: Berechnung von Quantilen: Sei $Z \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$ standardnormalverteilt. Für welchen Wert z gilt $\Pr(|Z| > z) = 5\%$?



Wegen der Symmetrie bzgl der y-Achse gilt

$$\Pr(|Z| > z) = \Pr(Z < -z) + \Pr(Z > z) = 2 \cdot \Pr(Z < -z)$$

Finde also $z > 0$, so dass $\Pr(Z < -z) = 2.5\%$. `> qnorm(0.025,mean=0,sd=1) [1] -1.959964` Antwort: $z \approx 1.96$, also knapp 2 Standardabweichungen

8 Normalapproximation

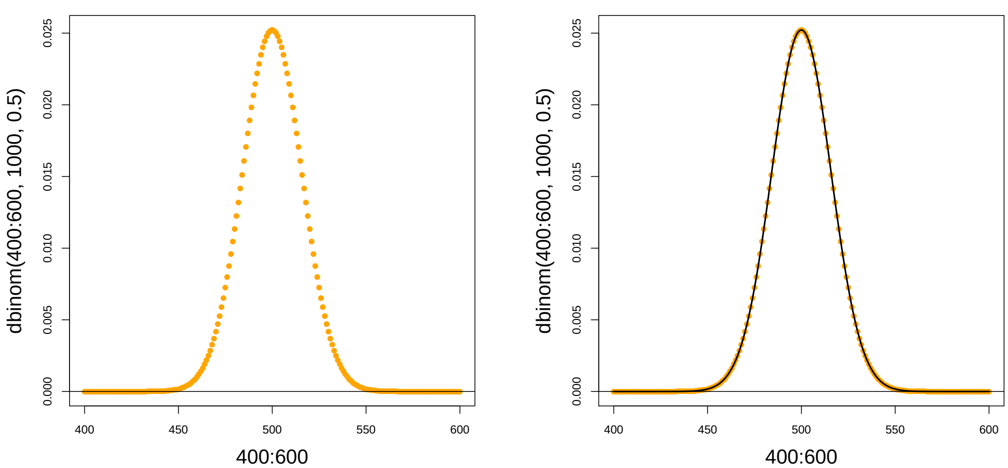
Normalapproximation

Für große n und p, die nicht zu nahe bei 0 oder 1 liegen, kann man die Binomialverteilung durch die Normalverteilung mit dem entsprechenden Erwartungswert und der entsprechenden Varianz approximieren:

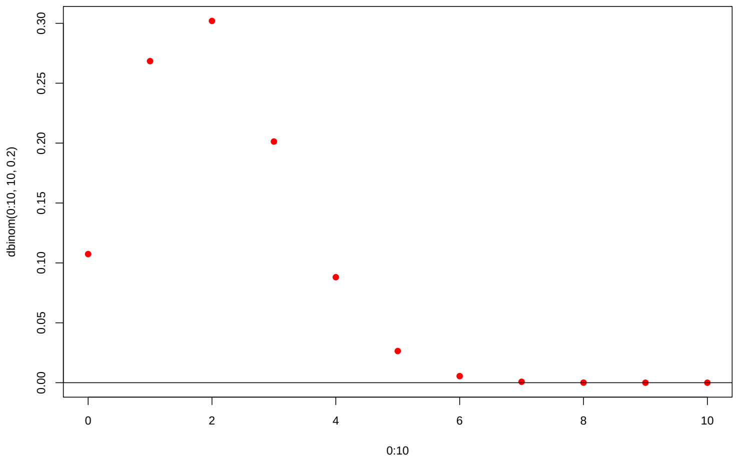
Ist $X \sim \text{bin}(n, p)$ und $Z \sim \mathcal{N}(\mu = n \cdot p, \sigma^2 = n \cdot p \cdot (1 - p))$, so gilt

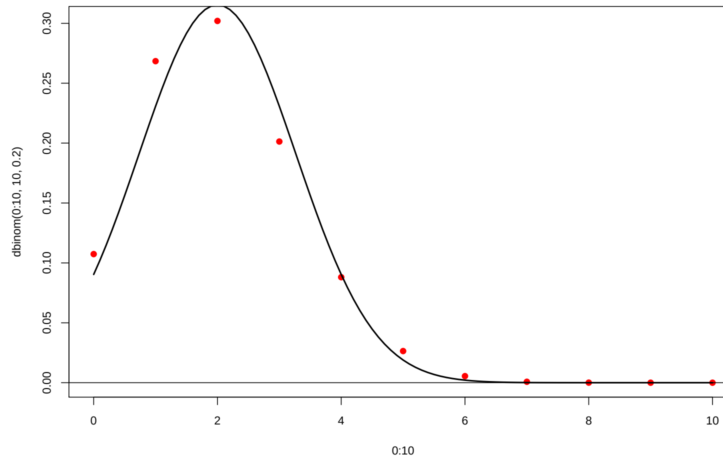
$$\Pr(X \in [a, b]) \approx \Pr(Z \in [a, b])$$

(eine Faustregel: für den Hausgebrauch meist okay, wenn $n \cdot p \cdot (1 - p) \geq 9$)
$n = 1000$, $p = 0.5$, $n \cdot p \cdot (1 - p) = 250$



$n = 10$, $p = 0.2$, $n \cdot p \cdot (1 - p) = 1.6$





Zentraler Grenzwertsatz

Eine etwas allgemeinere *Normalapproximation* beschreibt der Zentraler Grenzwertsatz.

Der zentrale Grenzwertsatz besagt, dass die Verteilung von Summen
unabhängiger und identisch verteilter
Zufallsvariablen in etwa die Normalverteilung ist.

Theorem 1 (Zentraler Grenzwertsatz) *Die $\mathbb{R}$-wertigen Zufallsgrößen $X_1, X_2, \ldots$ seien unabhängig und identisch verteilt mit endlicher Varianz $0 < Var\ X_i < \infty$. Sei außerdem*

$$Z_n := X_1 + X_2 + \cdots + X_n$$

die Summe der ersten n Variablen. Dann ist die zentrierte und reskalierte Summe im Limes $n \to \infty$ standardnormalverteilt, d.h.

$$\frac{Z_n - \mathbb{E}Z_n}{\sqrt{Var\ Z_n}} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$$

bei $n \to \infty$. Formal: Es gilt für alle $-\infty \leq a < b \leq \infty$

$$\lim_{n\to\infty} \Pr\left(a \leq \frac{Z_n - \mathbb{E}Z_n}{\sqrt{Var\ Z_n}} \leq b\right) = \Pr(a \leq Z \leq b),$$

wobei Z eine standardnormalverteilte Zufallsvariable ist.

Anders formuliert: Für große n gilt

$$Z_n \sim \mathcal{N}\big(\mu = \mathbb{E}Z_n, \sigma^2 = \mathrm{Var}\ Z_n\big)$$

Die Voraussetzungen „unabhängig“ und „identisch verteilt“ lassen sich noch deutlich abschwächen.

Für den Hausgebrauch:

Ist Y das Resultat von vielen kleinen Beiträgen, die großteils unabhängig voneinander sind, so ist Y in etwa normalverteilt,

d.h.

$$Y \sim \mathcal{N}\big(\mu = \mathbb{E}Y, \sigma^2 = \mathrm{Var}\ Y\big)$$

9 Der z-Test

Zurück zu dem Beispiel mit den Prolin-Codons im menschnlichen Genom.

CCT kommt $k = 16710$ mal vor CCC kommt $n - k = 18895$ mal vor

Frage: Kann das Zufall sein?
Wir meinen: Nein.

Die Skeptiker sagen: „Nur Zufall."

Die Hypothese
Reiner Zufall Kein Unterschied
nennt man die Nullhypothese.
Um die Skeptiker zu überzeugen, müssen wir die Nullhypothese entkräften d.h. zeigen, dass unter der Nullhypothese die Beobachtung sehr unwahrscheinlich ist.

CCT kommt $k = 16710$ mal vor CCC kommt $n - k = 18895$ mal vor Unter der Nullhypothese „alles nur Zufall" ist die Anzahl X der CCT $\text{bin}(n, p)$-verteilt mit $n = 35605$ und $p = 0.5$.

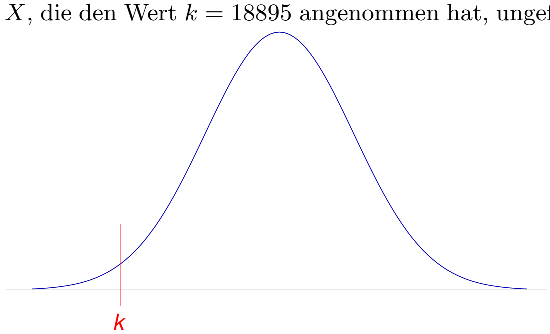
Normalapproximation: X ist ungefähr $\mathcal{N}(\mu, \sigma^2)$-verteilt mit

$$\mu = n \cdot p = 17802.5 \approx 17800$$

und

$$\sigma = \sqrt{n \cdot p \cdot (1-p)} = 94.34 \approx 95$$

Frage: Ist es plausibel, dass eine Größe X, die den Wert $k = 18895$ angenommen hat, ungefähr $\mathcal{N}(17800, 95^2)$-verteilt ist?

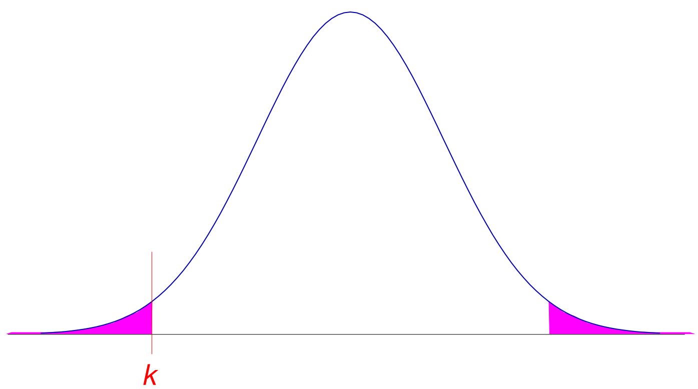


Wenn diese Nullhypothese H_0 gilt, dann folgt

$$\Pr(X = 17800) = 0$$

Aber das bedeutet nichts, denn $\Pr(X = k) = 0$ gilt für jeden Wert k!

Entscheidend ist die Wahrscheinlichkeit, dass X (unter Annahme der H_0) einen mindestens so extremen Wert wie k annimmt:



$$\Pr(|X-\mu| \geq |k-\mu|) = \Pr(|X-\mu| \geq 1092.5) \approx \Pr(|X-\mu| \geq 11.6 \cdot \sigma)$$

Wir wissen bereits:

$$Pr(|X-\mu| \geq 3 \cdot \sigma) \approx 0.003 \qquad \text{(siehe Merkregeln!)}$$

Also muss $\Pr(|X-\mu| \geq 11.6 \cdot \sigma)$ extrem klein sein.
In der Tat:

```
> 2 * pnorm(18895,mean=17800,sd=95,lower.tail=FALSE) [1] 9.721555e-31
```

Ohne Normalapproximation:

```
> pbinom(16710,size=35605,p=0.5) + + pbinom(18895-1,size=35605,p=0.5,lower.tail=FALSE) [1] 5.329252e-31
```

Wir können also argumentieren, dass eine derartig starke Abweichung vom Erwartungswert nur durch einen extremen Zufall zu erklären ist.

Wir werden also die Nullhypothese "alles nur Zufall" verwerfen und nach alternativen Erklärungen suchen, etwa unterschiedliche Effizienz von CCC und CCT oder unterschiedliche Verfügbarkeit von C und T.

Zusammenfassung z-Test

Nullhypothese H_0 (möchte man meistens verwerfen): der beobachtete Wert x kommt aus einer Normalverteilung mit Mittelwert μ und **bekannter** Varianz σ^2.

p-Wert $=\Pr(|X-\mu| \geq |x-\mu|)$, wobei $X \sim \mathcal{N}(\mu, \sigma^2)$, also die Wahrscheinlichkeit einer *mindestens* so großen Abweichung wie der beobachteten.

Signifikanzniveau α : oft 0.05. Wenn der p-Wert kleiner ist als α, verwerfen wir die Nullhypothese auf dem Signifikanzniveau α und suchen nach einer alternativen Erklärung.

Grenzen des z-Tests

Der z-Test kann nur angewendet werden, wenn die Varianz der Normalverteilung bekannt ist oder zumindest in der Nullhypothese als bekannt angenommen wird.

Das ist meistens nicht der Fall, wenn die Normalverteilung beim statistischen Testen verwendet wird.

Meistens wird die Varianz aus den Daten geschätzt. Dann muss statt dem z-Test der berühmte

t-Test

angewendet werden.

Was Sie u.a. erklären können sollten

- Wahrscheinlichkeitsdichten und was sie mit Wahrscheinlichkeitsverteilungen zu tun haben
- Wann und wie man die Binomialverteilung durch die Normalverteilung approximieren kann
- Eigenschaften der Normalverteilung (μ, σ^2, wichtige Quantile,...)
- Ist X normalverteilt, so ist auch $a \cdot X + b$ normalverteilt.
- Bedeutung des zentralen Grenzwertsatzes
- R-Befehle für Wahrscheinlichkeitsverteilungen
- z-Test, H_0, p-Wert, Signifikanzniveau α

Bitte beachten Sie auch die Auflistungen auf Seite 10 und Seite 21.