Aufgabe 1 Sehr viele Anwendungen wie z.B. Monte-Carlo-Simulationen benötigen stetige gleichverteilte Zufallszahlen. Entsprechend gibt es verschiedene Algorithmen um gleichverteilte Pseudozufallsvariablen zu generieren. Der Lehmer Algorithmus z.B., generiert rekursive Zufallszahlen mit der Formel

$$X_{k+1} = a \cdot X_k \mod m\,.$$

Im Jupyter-Notebook `Start_Hilfe.ipynb` auf Moodle finden Sie eine Implementierung; es werden mit dem Lehmer Algorithmus 100 Zahlen zwischen 0 und 1 generiert.

Überprüfen Sie rein visuell mit Hilfe von deskriptiver Statistik, ob die generierten Zahlen tatsächlich uniformverteilt sind oder nicht, (d.h. machen Sie einen geeigneten Histogram, einen Plot für ECDF und einen geeigneten Q-Q-Plot).

Machen Sie nun eine ähnliche Überprüfung für Zufallszahlen, die mit Numpy generiert werden. Mit Numpy können 100 gleichverteilte Zufallszahlen zwischen 0 und 1 so generiert werden

```
> np.random.uniform(0, 1, size = 100)
```

Kommentar: wichtig für die Prüfung
Lösen Sie die drei folgenden Aufgaben sowohl mit Python als auch mit der Tabelle für die Standardnormalverteilung (am Ende dieser Serie und auf Moodle zu finden).

Aufgabe 2 Eine Zufallsvariable Z heisst standardnormalverteilt, wenn ihre Verteilung durch folgende Dichtefunktion φ beschrieben wird:

$$\varphi(z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \mathbb{R}.$$

Diese Funktion nennt man auch die *Gausssche Glockenkurve*. Die zugehörige kumulative Verteilungsfunktion bezeichnen wir mit Φ:

$$\Phi(z) := \int_{-\infty}^{z} \varphi(t)dt.$$

Leider gibt es für diese Verteilungsfunktion keine geschlossene Formel! Die Standardnormalverteilung hat folgende charakteristische Eigenschaften:

- Ist Z standardnormalverteilt, so gilt:

  $$E(Z) = 0, \quad Var(Z) = \sigma(Z) = 1.$$

  Man kann auch leicht nachrechnen, dass der Graph der Dichtefunktion seine zwei Wendepunkte genau bei $\pm\sigma(Z) = \pm 1$ hat, siehe Abbildung 1.

- Die Dichtefunktion φ ist eine gerade Funktion, d.h. es gilt $\varphi(-z) = \varphi(z)$ für alle $z \in \mathbb{R}$. Für die Verteilungsfunktion Φ folgt damit, dass $\Phi(-z) = 1 - \Phi(z)$ für alle $z \in \mathbb{R}$ gilt (siehe Abbildung 1).

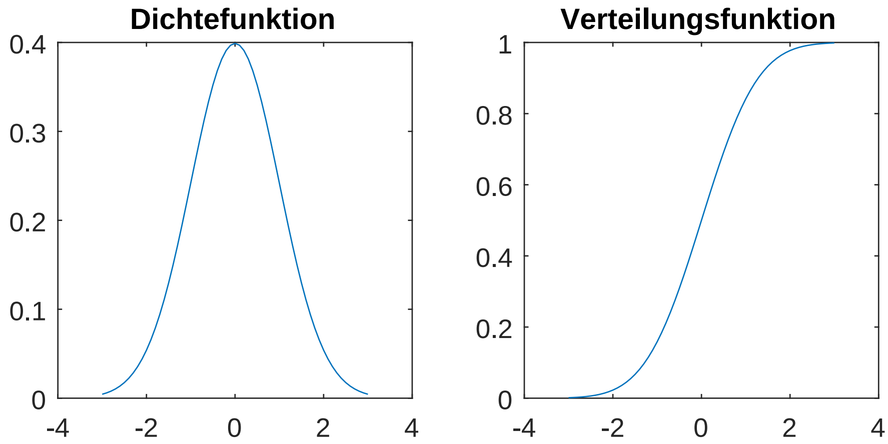


Abbildung 1 – Dichte- und Verteilungsfunktion der Standardnormalverteilung

- Wie bereits erwähnt, gibt es für die Verteilungsfunktion Φ keine geschlossene Formel. Die entsprechenden Werte erhält man aber wie für jede andere relevante Verteilung mit dem Python-Scipy-Befehl `cdf`: Die Wahrscheinlichkeit, dass eine standardnormalverteilte Zufallsvariable höchstens den Wert 2 annimmt, ist beispielsweise

  ```
  > from scipy import stats
  > stats.norm.cdf(2, loc = 0, scale = 1)
  ans = 0.9772498680518208
  ```

  Wir wählen den gewünschten Mittelwert mit Hilfe des zweiten Argumentes `loc` und die gewünschte Standardabweichung mit dem dritten Argument `scale`. Per Default wird so wieso die Standardnormalverteilung verwendet, also müssen wir diese Parameter hier nicht explizit angeben.

  Quantile können mit der sogenannten "percent point function" berechnet werden. Wir verwenden den Python-Scipy-Befehl `ppf` folgendermassen:

  ```
  > stats.norm.ppf(0.9772498680518208)
  ans = 2.0000000000000004
  ```

  Das wir nicht exakt 2 erhalten liegt an Rundungsfehlern (maschine precision).

Drücken Sie die folgenden Wahrscheinlichkeiten einer standardnormalverteilten Zufallsvariable Z mit Hilfe der Funktion Φ aus, und berechnen Sie die Wahrscheinlichkeiten mit Python und von Hand mit der Tabelle für Normalverteilung (auf Moodle zu finden).

(i) $P(Z \leq 1)$

(ii) $P(-1 \leq Z \leq 1)$

(iii) $P(|Z| \geq 0.5)$

(iv) $P(-3 \leq Z \leq 1)$

Berechnen Sie die folgenden Quantile in einem symmetrischen Interval um 0 herum.

(i) $P(x_1 \leq Z \leq x_2) \approx 0.6827$

(ii) $P(x_1 \leq Z \leq x_2) \approx 0.9545$

(iii) $P(x_1 \leq Z \leq x_2) \approx 0.9973$

Aufgabe 3 Wir betrachten nun Normalverteilungen mit beliebigem Erwartungswert und beliebiger Varianz. Eine Zufallsvariable heisst normalverteilt mit Mittelwert $\mu \in \mathbb{R}$ und Standardabweichung $\sigma > 0$, wenn sie sich schreiben lässt als

$$X = \mu + \sigma Z,$$

wobei Z standardnormalverteilt ist. Für die so definierte Zufallsvariable X gilt tatsächlich

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

Die Dichtefunktion von X ist gegeben durch

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

und die kumulative Verteilungsfunktion lässt sich schreiben als

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Die Dichtefunktion von X ist also eine „Gaussglocke" mit Zentrum bei μ und Standardabweichung σ, wobei man diese Standardabweichung graphisch wieder an den Wendepunkten erkennen kann. Abbildung 2 illustriert dies im Falle von $\mu = 3$ und $\sigma = 0.5$.

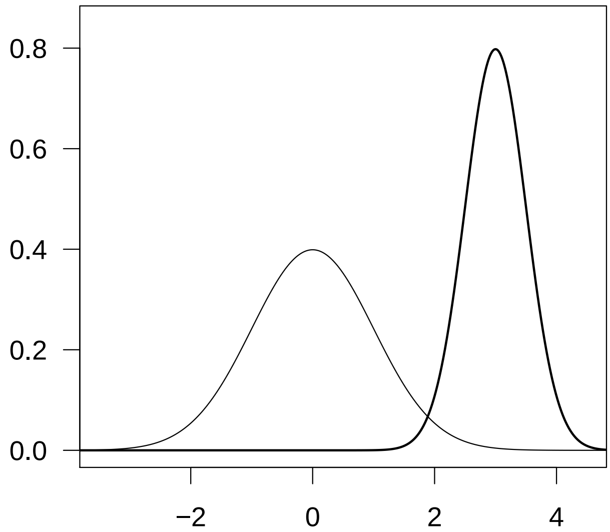


Abbildung 2 – Dichtefunktion der Standardnormalverteilung und einer Normalverteilung mit $\mu = 3$ und $\sigma = 0.5$

Umgekehrt können wir eine normalverteilte Zufallsvariable X immer mit Hilfe der Transformation

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

auf eine standardnormalverteilte Zufallsvariable zurückführen (vgl. vorherige Aufgabe). Manchmal spricht man auch vom „z-Wert" einer Messung x_i und meint damit den Wert $z_i = \frac{x_i-\mu}{\sigma}$, also den Abstand vom Erwartungswert μ (einer angenommenen/geschätzten Normalverteilung) *ausgedrückt als Anzahl Standardabweichungen* σ (der angenommenen/geschätzten Normalverteilung).

In diesem Zusammenhang erwähnen wir auch die folgenden Faustregeln, die Sie auch schon angetroffen haben: Bei normalverteilten Messungen liegen

- ca. 68% der Messwerte innerhalb *einer* Standardabweichung vom Mittelwert (haben also einen z-Wert zwischen -1 und 1)

- ca. 95% der Messwerte innerhalb von *zwei* Standardabweichungen vom Mittelwert.

- ca. 99.7% der Messwerte innerhalb von *drei* Standardabweichungen vom Mittelwert.

Drücken Sie die folgenden Wahrscheinlichkeiten mit Hilfe der Funktion Φ aus, und berechnen Sie die Wahrscheinlichkeiten mit Python und mit der Tabelle für die Normalverteilung.

(i) $P(X \leq 2)$, wobei X normalverteilt mit $\mu = 1$ und $\sigma^2 = 9$ ist, d.h. $X \sim \mathcal{N}(\mu, \sigma^2)$.

(ii) $P(-4 \leq X \leq 8)$, wobei X normalverteilt mit $\mu = 2$ und $\sigma = 2$ ist. Das wird vielfach auch so geschrieben $X \sim \mathcal{N}(2, 4)$.

(iii) $P(|X| \leq 1)$, wobei X normalverteilt mit $\mu = -11$ und $\sigma^2 = 16$ ist. Das wird vielfach so geschrieben $X \sim \mathcal{N}(-11, 4^2)$.

Aufgabe 4 Der Body-Mass-Index (BMI) einer Person ist folgendermassen definiert:

$$\text{BMI} = \frac{\text{Masse in kg}}{(\text{Körperlänge in m})^2}$$

Wir nehmen an, dass der BMI in den USA durch die Zufallsvariable $X \sim \mathcal{N}(27, 6^2)$ beschrieben wird. Das nationale Herz-, Lungen- und Blutinstitut definiert die folgenden BMI-Kategorien:

- Untergewicht: BMI ≤ 18.5

- Normalgewicht: $18.5 <$ BMI ≤ 25

- Übergewicht: $25 <$ BMI ≤ 30

- Fettleibigkeit: BMI > 30

(a) Welcher Prozentsatz der Bevölkerung hat Übergewicht?

(b) Welcher Prozentsatz hat Untergewicht oder leidet an Fettleibigkeit?

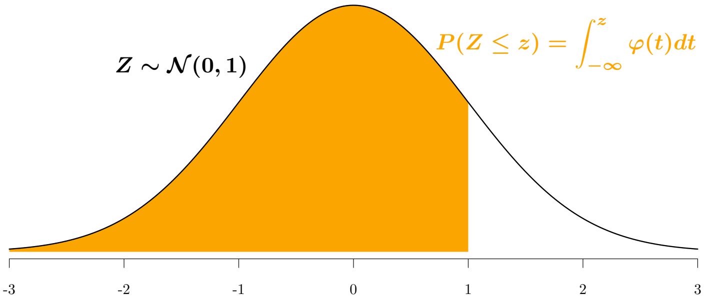
(c) 5% der Bevölkerung haben einen BMI, der grösser als x ist. Bestimmen Sie x.

Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Area under the Standard Normal Curve

by Vidushi Bigler

This table provides the area between $-\infty$ and some positiv Z score under the standard normal curve. For example, when Z score is equal to 1, then the area is 0.8413. For Z score equal to -1, the area is $1 - 0.8413 = 0.1587$. When the Z score is 1.45, the area is 0.9265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990