Die **Verteilungsfunktion** F einer Zufallsvariablen X ordnet jeder reellen Zahl x die Wahrscheinlichkeit

$$F(x) = P(X \leq x)$$

zu.

<u>Beispiel:</u> X hat eine Bernoulli-Verteilung mit Erfolgswahrscheinlichkeit p=0.3 und nimmt daher nur zwei Werte an, den Wert 1 (Erfolg) mit der Wahrscheinlichkeit

$$P(X = 1) = p = 0.3$$

und den Wert 0 (Misserfolg) mit der Wahrscheinlichkeit

$$P(X = 0) = 1 - p = 0.7.$$

$$F(-0.1) = P(X \leq -0.1) = 0$$

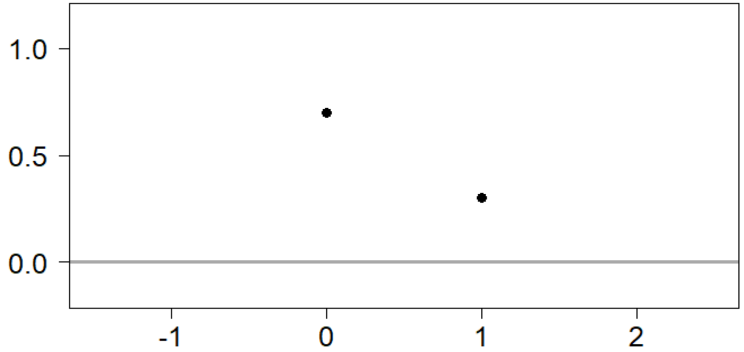
$$F(0) = P(X \leq 0) = 0.7$$

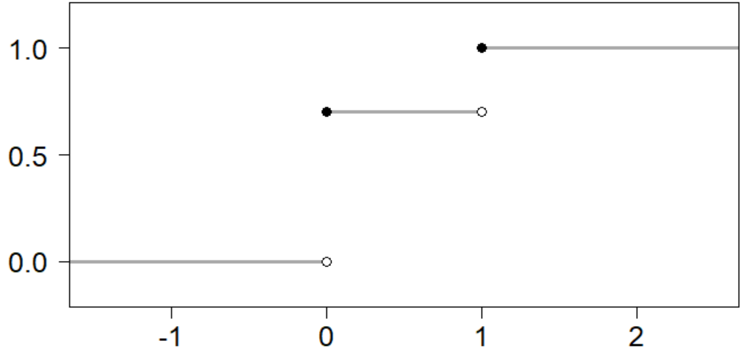
$$F(0.5) = P(X \leq 0.5) = 0.7$$

$$F(1) = P(X \leq 1) = 1$$

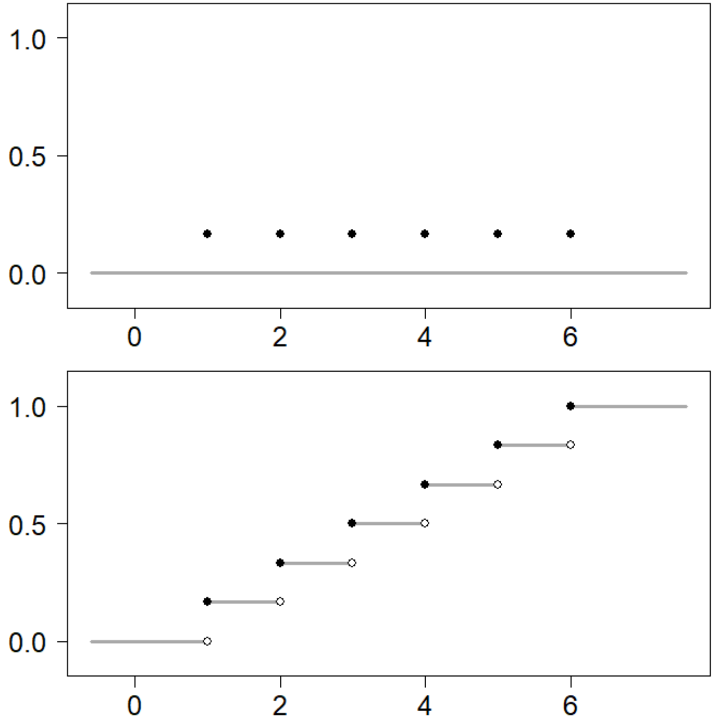
$$F(1.5) = P(X \leq 1.5) = 1$$

Wahrscheinlichkeiten und Verteilungsfunktion:





<u>Beispiel:</u> Gleichverteilung mit Träger {1,2,3,4,5,6}



Zufallsvariablen, deren Verteilungsfunktionen durch Treppenfunktionen gegeben sind, heißen diskret. Beispielsweise sind Bernoulli-verteilte, binomialverteilte und gleichverteilte Zufallsvariablen diskret.

Erhöht man die Anzahl der möglichen Werte einer Gleichverteilung, dann verringert sich die Sprunghöhe entsprechend. Asymptotisch verschwindet sie gänzlich, man erhält also eine stetige Verteilungsfunktion.

Eine Zufallsvariable X heißt **stetig**, wenn eine Funktion f: $\mathbf{R}\to[0,\infty)$ existiert, sodass für alle Borel-Mengen A gilt:

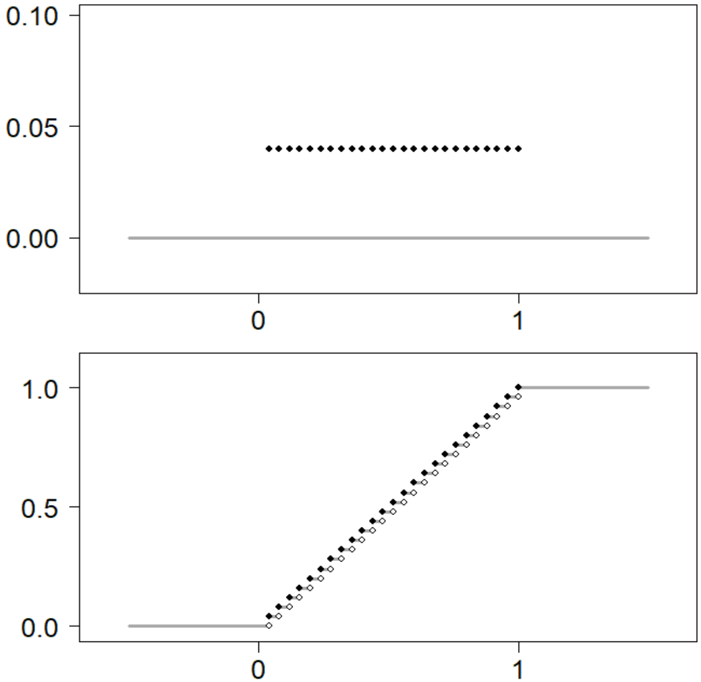
$$P(X \in A) = \int_A f(t)\,dt$$

Eine solche Funktion f heißt **Dichtefunktion** von X.
Die zugehörige Verteilungsfunktion ist gegeben durch

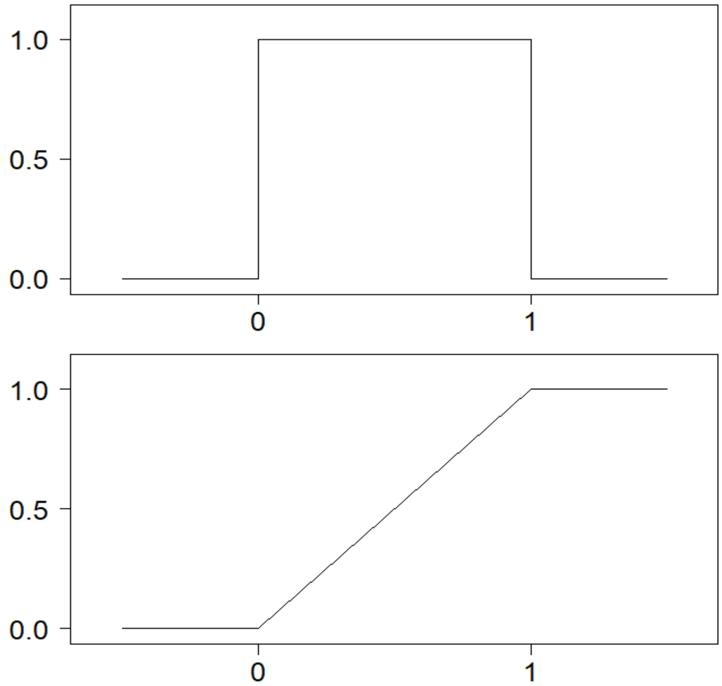
$$F(x) = P(X \le x) = P(X \in (-\infty, x]) = \int_{(-\infty,x]} f(t)\,dt = \int_{-\infty}^{x} f(t)\,dt$$

und ihre Ableitung F' durch f.

<u>Beispiel:</u> Gleichverteilung mit Träger {1/25,2/25,…,1}



<u>Beispiel:</u> Dichtefunktion und Verteilungsfunktion der stetigen Gleichverteilung auf [0,1]



Beispiel:

X gleichverteilt auf [−0.5,1.5], $P(X<0.1)=?$

Lösung:

Die Zufallsvariable X hat eine stetige Gleichverteilung auf dem Intervall [−0.5,1.5]. Auf diesem Intervall ist ihre Dichte f also konstant und verschwindet außerhalb:

$$f(x) = \begin{cases} 0, & x < -0.5 \\ c, & -0.5 \leq x \leq 1.5 \\ 0, & x > 1.5 \end{cases}$$

Weiters folgt aus

$$1 = P\big(X \in (-\infty, \infty)\big) = \int_{-\infty}^{\infty} f(x)dx$$

$$= \int_{-\infty}^{-0.5} f(x)dx + \int_{-0.5}^{1.5} f(x)dx + \int_{1.5}^{\infty} f(x)dx$$

$$= \int_{-\infty}^{-0.5} 0\, dx + \int_{-0.5}^{1.5} c\, dx + \int_{1.5}^{\infty} 0\, dx$$

$$= c\int_{-0.5}^{1.5} 1\, dx = c\big(1.5 - (-0.5)\big) = 2c,$$

dass

$$c = \frac{1}{2}\,.$$

Analog erhält man:

$$P\big(X < 0.1\big) = P\big(X \in (-\infty, 0.1)\big) = \int_{-\infty}^{0.1} f(x)dx$$

$$= \int_{-\infty}^{-0.5} f(x)dx + \int_{-0.5}^{0.1} f(x)dx$$

$$= \int_{-\infty}^{-0.5} 0\, dx + \int_{-0.5}^{0.1} \frac{1}{2}\, dx$$

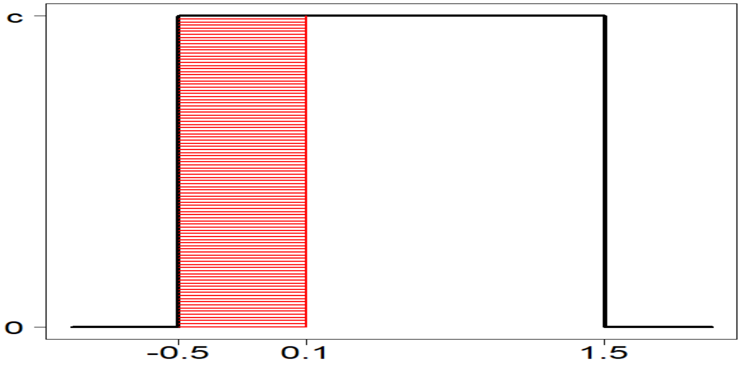
$$= \frac{1}{2}\big(0.1 - (-0.5)\big)$$

$$= \mathbf{0.3},$$

Beispiel:

X gleichverteilt auf $[-0.5, 1.5]$, $P(X<0.1)=?$

Schnelle Lösung mit Skizze:



Gesamtfläche unter Dichte = Fläche des großen Rechtecks

$$= c\,(1.5-(-0.5)) = 1$$

$$\Rightarrow c = 0.5$$

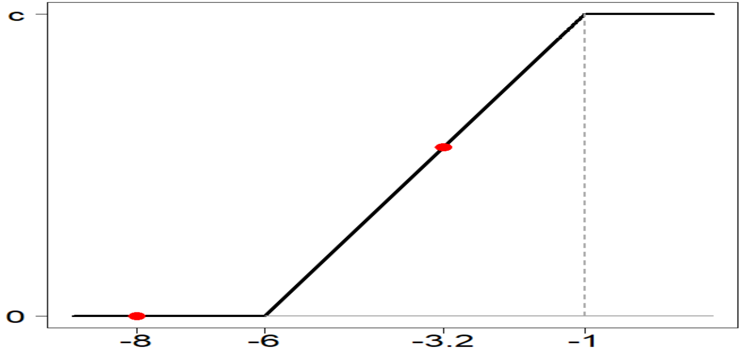
$P(X<0.1)$ = Fläche des roten Rechtecks

$$= 0.5\,(0.1-(-0.5)) = \mathbf{0.3}$$

Beispiel:

Verteilungsfunktion von X für $-6<x<-1$: $F(x)=1.2+0.2x$, $P(-8<X<-3.2)=?$

Schnelle Lösung mit Skizze:



$$P(-8<X<-3.2) = \int_{-8}^{-3.2} f(x)dx = F(-3.2) - F(-8)$$

$$= \big(1.2 + 0.2\cdot(-3.2)\big) - 0$$

$$= \mathbf{0.56}$$

Das arithmetische Mittel

$$\bar{X} = \frac{X_1 + X_2}{2}$$

von zwei unabhängigen Zufallsvariablen X_1 und X_2, die beide Bernoulli-verteilt sind mit derselben Erfolgswahrscheinlichkeitk p, ist nicht Bernoulli-verteilt, weil es mehr als zwei Werte annehmen kann.

$$X_1 = 0,\ X_2 = 0 \Rightarrow \bar{X} = 0,$$
$$X_1 = 0,\ X_2 = 1 \vee X_1 = 1,\ X_2 = 0 \Rightarrow \bar{X} = ½,$$
$$X_1 = 1,\ X_2 = 1 \Rightarrow \bar{X} = 1.$$

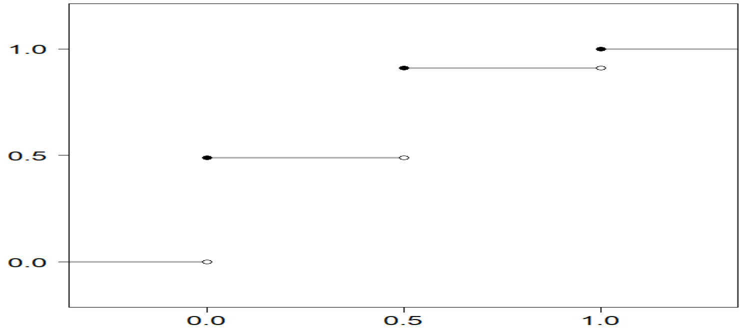
Wahrscheinlichkeiten der drei möglichen Werte 0, ½ und 1:

$$P(\bar{X} = 0) = P(X_1 + X_2 = 0) = (1-p)^2$$

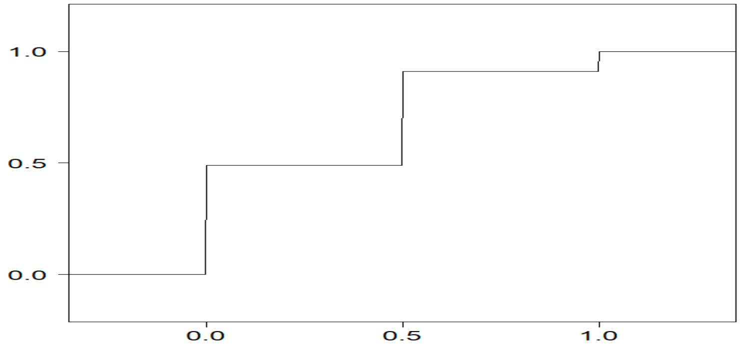
$$P(\bar{X} = ½) = P(X_1 + X_2 = 1)$$
$$= P\big((X_1 = 0 \wedge X_2 = 1) \vee (X_1 = 1 \wedge X_2 = 0)\big)$$
$$= 2p(1-p)$$

$$P(\bar{X} = 1) = P(X_1 + X_2 = 2) = p^2$$

Verteilungsfunktion von $\bar{X}$ für den Fall $p = 0.3$



Vereinfachte Darstellung

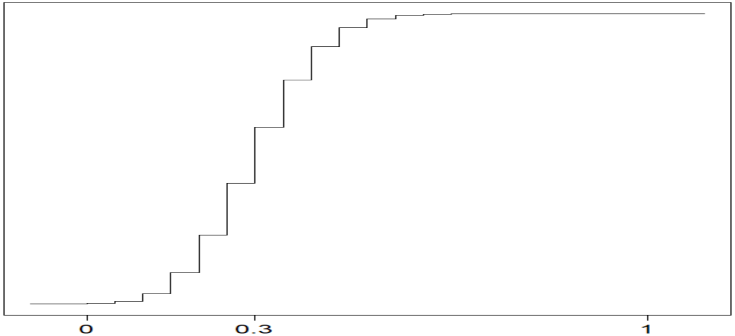


Die Verteilungsfunktion des arithmetischen Mittels

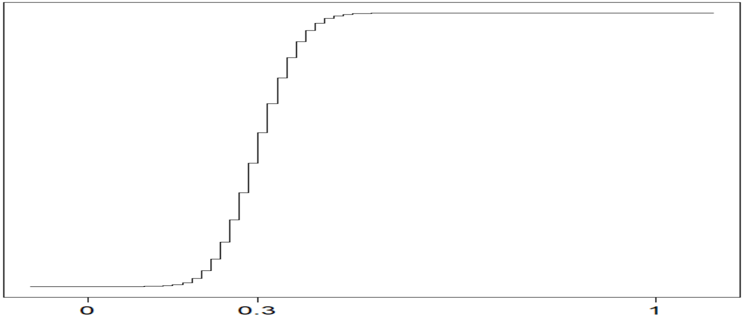
$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

von n unabhängigen und identisch Bernoulli-verteilten Zufallsvariablen mit Parameter p nähert sich mit wachsendem n einer Treppenfunktion an, die nur einen Sprung an der Stelle p der Höhe 1 hat. Eine solche Funktion ist die Verteilungsfunktion einer Zufallsvariablen, die nur einen einzigen Wert annehmen kann.

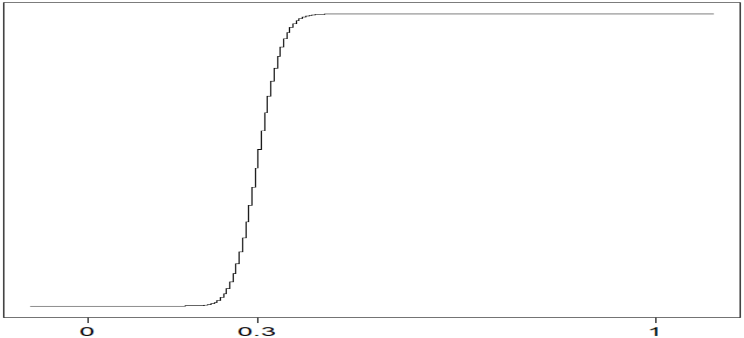
Verteilungsfunktion von $\bar{X}$ für n=20, p=0.3



Verteilungsfunktion von $\bar{X}$ für n=60, p=0.3



Verteilungsfunktion von $\bar{X}$ für n=180, p=0.3



Dass die Verteilung des arithmetischen Mittels $\bar{X}$ sich mit wachsendem n auf nur einen Punkt, nämlich den Erwartungswert, konzentriert, gilt nicht nur für die Bernoulli-Verteilung, sondern auch für viele andere Verteilungen (**Gesetz der großen Zahlen**).

Das liegt daran, dass der Erwartungswert von $\bar{X}$ mit wachsendem n konstant bleibt, die Varianz von $\bar{X}$ aber verschwindet.

Sind $X_1, \dots, X_n$ unabhängig und identisch verteilt (i.i.d.) mit Erwartungswert μ und Varianz σ^2, dann gilt

$$\begin{aligned} E(\bar{X}) &= E(\tfrac{1}{n}(X_1 + \dots + X_n)) \\ &= \tfrac{1}{n} E(X_1 + \dots + X_n) \\ &= \tfrac{1}{n}(E(X_1) + \dots + E(X_n)) \\ &= \tfrac{1}{n}(\mu + \dots + \mu) \\ &= \tfrac{1}{n} n\mu \\ &= \mu \end{aligned}$$

und

$$\begin{aligned} \operatorname{var}(\bar{X}) &= \operatorname{var}(\tfrac{1}{n}(X_1 + \dots + X_n)) \\ &= \tfrac{1}{n^2} \operatorname{var}(X_1 + \dots + X_n) \\ &= \tfrac{1}{n^2}(\operatorname{var}(X_1) + \dots + \operatorname{var}(X_n)) \\ &= \tfrac{1}{n^2}(\sigma^2 + \dots + \sigma^2) \\ &= \tfrac{1}{n^2} n\sigma^2 \\ &= \tfrac{\sigma^2}{n} \to 0, \text{ wenn } n \to \infty. \end{aligned}$$

Die Varianz des arithmetischen Mittels ist gegeben durch

$$var(\bar{X}) = \frac{\sigma^2}{n}$$

und verschwindet daher, wenn $n \to \infty$. Wenn man allerdings $\bar{X}$ mit $\sqrt{n}$ multipliziert, so erhält man asymptotisch zwar eine nichtverschwindende Varinaz,

$$var\left(\sqrt{n}\bar{X}\right) = \left(\sqrt{n}\right)^2 \frac{\sigma^2}{n} = \sigma^2,$$

aber dafür explodiert dann der Erwartungswert,

$$E\left(\sqrt{n}\bar{X}\right) = \sqrt{n}E(\bar{X}) = \sqrt{n}\mu \to \infty.$$

Das kann vermieden werden, wenn man statt der ursprünglichen Zufallsvariablen $X_1, \dots, X_n$ die zentrierten Zufallsvariablen $X_1 - \mu, \dots, X_n - \mu$ oder die standardisierten Zufallsvariablen

$$Z_i = \frac{X_i - \mu}{\sigma}, \; i=1,\dots,n,$$

betrachtet.

Die standardisierten Zufallsvariablen Z_i haben Erwartungswert 0 und Varianz 1:

$$E(Z_i) = E\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sigma} E(X_i - \mu)$$
$$= \frac{1}{\sigma}\left(E(X_i) - E(\mu)\right) = \frac{1}{\sigma}(\mu - \mu) = 0,$$

$$var(Z_i) = var\left(\frac{X_i - \mu}{\sigma}\right) = \frac{1}{\sigma^2} var(X_i - \mu)$$
$$= \frac{1}{\sigma^2} var(X_i) = \frac{1}{\sigma^2}\sigma^2 = 1$$

Daraus folgt, dass

$$E\left(\sqrt{n}\bar{Z}\right) = \sqrt{n}E(\bar{Z}) = \sqrt{n} \cdot 0 = 0$$

und

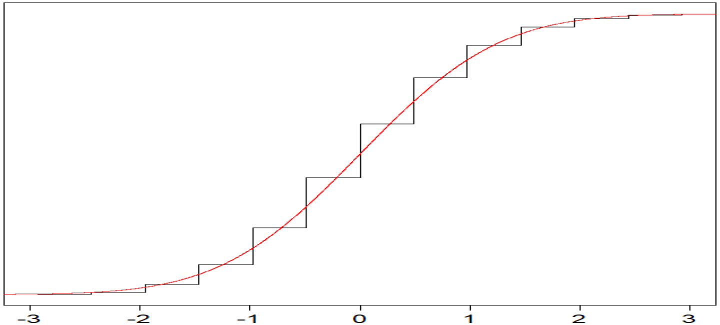
$$var\left(\sqrt{n}\bar{Z}\right) = \left(\sqrt{n}\right)^2 var(\bar{Z}) = n \cdot \frac{1}{n} = 1.$$

Im Fall von standardisierten i.i.d. Bernoulli-verteilten Zufallsvariablen mit p=0.3 nähert sich die Verteilungsfunktion von $\sqrt{n}\bar{Z}$ der stetigen Funktion

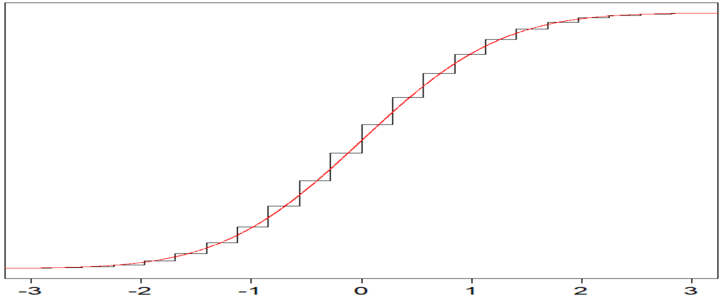
$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) \mathrm{dt}$$

an. Eine stetige Zufallsvariable, deren Verteilungsfunktion gegeben ist durch Φ, heißt standardnormalverteilt.

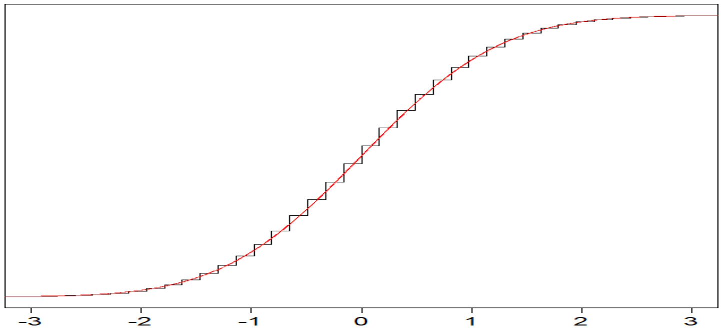
Verteilungsfunktion von $\sqrt{n}\bar{Z}$ für n=20, p=0.3



Verteilungsfunktion von $\sqrt{n}\bar{Z}$ für n=60, p=0.3



Verteilungsfunktion von $\sqrt{n}\bar{Z}$ für n=180, p=0.3



Im Unterschied zu Gesetzen der großen Zahlen (LLN), die Aussagen machen über die Konvergenz von $\bar{X}$ gegen die degenerierte Zufallsvariable μ, die Varianz 0 hat, machen **zentrale Grenzwertsätze** (CLT) Aussagen über die Konvergenz von $\sqrt{n}\bar{Z} = \sqrt{n}(\bar{X} - \mu)/\sigma$ gegen eine standardnormalverteilte Zufallsvariable mit Erwartungswert 0 und Varianz 1.

Allgemein ist der Erwartungswert einer stetigen Zufallsvariablen mit Dichtefunktion f gegeben durch

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

und die Varianz (wie schon im diskreten Fall) durch

$$var(X) = E\big(X - E(X)\big)^2.$$

Die Rechenregeln für Erwartungswerte und Varianzen im stetigen Fall sind analog zu denen im diskreten Fall.

Beispiel:

$X_1, X_2, \ldots$ i.i.d. Bernoulli(0.8), $\sqrt{n}(\bar{X}-0.8)/\sqrt{0.2} \rightarrow Y$, var($Y$)=?

Lösung:

Gegeben ist eine Folge unabhängiger und identisch Bernoulli-verteilter Zufallsvariablen $X_1, X_2, \ldots$ mit Erfolgswahrscheinlichkeit $p = 0.8$. Die standardisierten Zufallsvariablen sind wegen

$$E(X_i) = p,\ var(X_i) = p(1-p)$$

gegeben durch

$$Z_i = \frac{X_i - \mu}{\sigma} = \frac{X_i - p}{\sqrt{p(1-p)}} = \frac{X_i - 0.8}{\sqrt{0.8}\,\sqrt{0.2}}.$$

Zentraler Grenzwertsatz:

$$\sqrt{n}\bar{Z} = \sqrt{n}\frac{\bar{X}-0.8}{\sqrt{0.8}\,\sqrt{0.2}} = \frac{1}{\sqrt{0.8}}\sqrt{n}\frac{\bar{X}-0.8}{\sqrt{0.2}} \rightarrow Z,\ var(Z) = 1$$

$$\Rightarrow \sqrt{n}\frac{\bar{X}-0.8}{\sqrt{0.2}} = \sqrt{0.8}\,\sqrt{n}\bar{Z} \rightarrow \sqrt{0.8}\,Z = Y$$

$$\Rightarrow var(Y) = var\big(\sqrt{0.8}\,Z\big) = \big(\sqrt{0.8}\big)^2 var(Z) = \mathbf{0.8}$$

Mit Hilfe der Transformation

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

erhält man aus einer standardnormalverteilten Zufallsvariablen Z eine normalverteilte Zufallsvarible X, deren Erwartungswert und Varianz gegeben sind durch

$$E(X) = E(\mu) + E(\sigma Z) = \mu + \sigma E(Z) = \mu + \sigma \cdot 0 = \mu$$

und

$$var(X) = var(\sigma Z) = \sigma^2 E(Z) = \sigma^2 \cdot 1 = \sigma^2.$$

Quadriert man eine standardnormalverteilte Zufallsvariable Z, so erhält man eine χ^2-verteilte Zufallsvariable X mit einem Freiheitsgrad, die den Erwartungswert

$$E(X) = E(Z^2) = E(Z-0)^2 = E(Z-EZ)^2$$
$$= var(Z) = 1$$

und die Varianz 2 hat.

Summiert man k Zufallsvariablen $X_1, \ldots, X_k$, die unabhängig und identisch χ^2-verteilt mit einem Freiheitsgrad sind, so erhält man eine χ^2-verteilte Zufallsvariable X mit k Freiheitsgraden, die den Erwartungswert

$$E(X) = E(X_1 + \cdots + X_k) = E(X_1) + \cdots + (X_k)$$
$$= 1 + \cdots + 1 = k$$

und die Varianz

$$var(X) = var(X_1 + \cdots + X_k) = var(X_1) + \cdots + var(X_k)$$
$$= 2 + \cdots + 2 = 2k$$

hat.

Multipliziert man eine χ^2-verteilte Zufallsvariable X mit 2 Freiheitsgraden mit dem Faktor $\tau/2$, so erhält man eine exponentialverteilte Zufallsvariable Y mit dem Erwartungswert

$$E(Y) = E\left(\frac{\tau}{2}X\right) = \frac{\tau}{2}E(X) = \frac{\tau}{2} \cdot 2 = \tau$$

und der Varianz

$$var(Y) = var\left(\frac{\tau}{2}X\right) = \left(\frac{\tau}{2}\right)^2 var(X) = \frac{\tau^2}{4} \cdot (2 \cdot 2) = \tau^2.$$

Ist die standardnormalverteilte Zufallsvariable Z unabhängig von der mit k Freiheitsgraden χ^2-verteilten Zufallsvariablen X, dann ist die Zufallsvariable

$$T = \frac{Z}{\sqrt{\frac{X}{k}}}$$

t-verteilt mit k Freiheitsgraden und ihr Erwartungswert und ihre Varianz sind gegeben durch

$$E(T) = 0$$

und

$$var(Y) = \frac{k}{k-2}\,.$$

Abkürzungen:

$X \sim N(\mu,\sigma^2)$ … X ist normalverteilt mit Parametern μ, σ^2

$X \sim \chi^2(k)$ ……. X ist χ^2-verteilt mit k Freiheitsgraden

$X \sim t(k)$ ……… X ist t-verteilt mit k Freiheitsgraden

$X \sim Exp(\tau)$ …… X ist exponentialverteilt mit Parameter τ

Beispiel:

$X_1, X_2, \ldots$ i.i.d. $t(6)$, $\sqrt{n}\bar{X}/\sqrt{6} \to Y$, var($Y$)=?

Lösung:

Gegeben ist eine Folge unabhängiger und identisch t-verteilter Zufallsvariablen $X_1, X_2, \ldots$ mit 6 Freiheitsgraden. Die standardisierten Zufallsvariablen sind wegen

$$E(X_i) = 0,\ var(X_i) = \frac{k}{k-2}$$

gegeben durch

$$Z_i = \frac{X_i-\mu}{\sigma} = \frac{X_i-0}{\sqrt{\frac{k}{k-2}}} = \frac{\sqrt{k-2}}{\sqrt{k}} X_i = \frac{\sqrt{4}}{\sqrt{6}} X_i.$$

Zentraler Grenzwertsatz:

$$\sqrt{n}\bar{Z} = \sqrt{n}\frac{\sqrt{4}}{\sqrt{6}}\bar{X} = \sqrt{4}\,\sqrt{n}\frac{\bar{X}}{\sqrt{6}} \to Z,\ \ var(Z) = 1$$

$$\Rightarrow \sqrt{n}\frac{\bar{X}}{\sqrt{6}} = \frac{1}{\sqrt{4}}\,\sqrt{n}\bar{Z} \to \frac{1}{\sqrt{4}} Z = Y$$

$$\Rightarrow var(Y) = var\left(\frac{1}{\sqrt{4}}\,Z\right) = \left(\frac{1}{\sqrt{4}}\right)^2 var(Z) = \mathbf{0.25}$$

Beispiel:

$X_1, X_2, \ldots$ i.i.d. $\chi^2(1)$, $\sqrt{n}(\bar{X}-1)/\sqrt{2} \to Y$, var($Y$)=?

Lösung:

Gegeben ist eine Folge unabhängiger und identisch χ^2-verteilter Zufallsvariablen $X_1, X_2, \ldots$ mit k=1 Freiheitsgraden. Die standardisierten Zufallsvariablen sind wegen

$$E(X_i) = k = 1, var(X_i) = 2k = 2$$

gegeben durch

$$Z_i = \frac{X_i - \mu}{\sigma} = \frac{X_i - 1}{\sqrt{2}} = \frac{X_i - 1}{\sqrt{2}}.$$

Zentraler Grenzwertsatz:

$$\sqrt{n}\bar{Z} = \sqrt{n}\frac{\bar{X}-1}{\sqrt{2}} \to Y, \; var(Y) = \mathbf{1}$$

Beispiel:

X, Y, Z i.i.d. $N(a, b^2), a = 0, b^2 = 3, var(X^2+Y^2+Z^2)$=?

Lösung:

Wären X, Y, Z unabhängig und identisch standardnormalverteilt, dann wäre die Summe $X^2+Y^2+Z^2$ χ^2(3)-verteilt und ihre Varianz wäre gleich 2·3=6.

Leider sind diese Zufallsvariablen aber nicht standardnormalverteilt, weil sie zwar Erwartungswert 0 haben, aber nicht Varianz 1, sondern Varianz $b^2 = 3$.

Allerdings sind die standardisierten Variablen $X/\sqrt{3}$, $Y/\sqrt{3}$, $Z/\sqrt{3}$ standardnormalverteilt, woraus folgt, dass

$$\left(\frac{X}{\sqrt{3}}\right)^2 + \left(\frac{Y}{\sqrt{3}}\right)^2 + \left(\frac{Z}{\sqrt{3}}\right)^2 = \frac{X^2+Y^2+Z^2}{3} \sim \chi^2(3)$$

und weiters

$$var(X^2+Y^2+Z^2) = var\left(3\frac{X^2+Y^2+Z^2}{3}\right)$$

$$= 9\, var\left(\frac{X^2+Y^2+Z^2}{3}\right) = 9 \cdot (2 \cdot 3) = \mathbf{54}.$$

Beispiel:

X, Y unabhängig, $X \sim N(a, b^2)$, $Y \sim N(c, d^2)$, $a = -6$, $b^2 = 3$, $c = 7$, $d^2 = 11$, $var(7 + 4X - Y) = ?$

Lösung:

$$\begin{aligned} var(7 + 4X - Y) &= var(4X - Y) \\ &= var(4X) + var(-Y) \\ &= 4^2 var(X) + (-1)^2 var(Y) \\ &= 16\, b^2 + d^2 \\ &= \mathbf{59} \end{aligned}$$