Kenngrößen von Zufallsvariablen

Die Wahrscheinlichkeitsverteilung kann durch die sogenannten *Kenngrößen* beschrieben werden, sie charakterisieren sozusagen die Verteilung.

Der Erwartungswert

Der **Erwartungswert** einer Verteilung ist definiert als

- **Diskrete Verteilung**

  $$\boxed{E(X) = \mu = \sum_{i=1}^{n} x_i P(X = x_i)}$$

  D.h. der Erwartungswert ist die Summe über alle Ausprägungen mal ihrer Wahrscheinlichkeit

  Beispiel: Wir wollen den Erwartungswert einer (diskreten) Variablen berechnen, deren Verteilung die *Poissonverteilung* ist, d.h.

  $$P(X = x_i) = P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}.$$

  Der Erwartungswert ist

  $$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \sum_{i=1}^{\infty} \lambda \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} \underbrace{\sum_{i=0}^{\infty} \frac{\lambda^i}{(i)!}}_{=e^{\lambda}} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned}$$

- **Stetige Verteilung**

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

  Es wird also über den Raum aller möglichen Ausprägungen (der Integrand ist die Ausprägung mal der Dichtefunktion) integriert.

Beispiel: Nehmen wir an, die Zufallsvariable X ist standardnormalverteilt, d.h. $X \sim N(0,1)$. Der Erwartungswert von X ist demnach 0 (eine Verteilung ist durch ihre Parameter charakterisiert, im Falle der Normalverteilung von μ, dem Erwartungswert, und von σ, der Standardabweichung).

Der Median und andere Quantile

Der Lageparamter **Median** ist einfach ausgedrückt der mittlere Wert einer Verteilung.

Nehmen wir an, die Zufallsvariable X hat folgende Realisierungen: $1, 4, 7, 13, 15$. Der Median von X wäre dann 7.

Der Median (symbolisch $\xi_{0.5}$) ist jener Wert, der die folgenden Bedingungen erfüllt:

$$P(X \leq \xi_{0.5}) \geq 0.5 \quad \text{und} \quad P(X \geq \xi_{0.5}) \geq 0.5.$$

Falls man die Verteilungsfunktion $F_X(t)$ der Zufallsvariable X invertieren kann, so berechnet sich der Median durch

$$\xi_{0.5} = F_X(0.5)^{-1}$$

Der Median ist ein Spezialfall des sogenannten α Quantils:

$$\xi_\alpha = F_X^{-1}(\alpha)$$

Weitere Spezialfälle sind

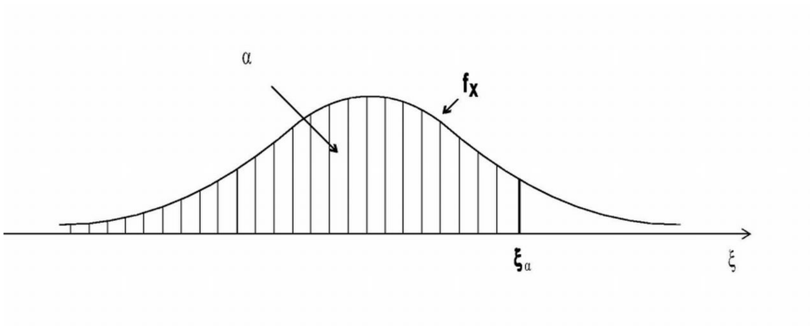
- **unteres Quartil**: 0.25 Quantil ($\xi_{0.25}$)
- **oberes Quartil**: 0.75 Quantil ($\xi_{0.75}$)

Wir unterscheiden wiederum den stetigen und den diskreten Fall:

- Falls die Zufallsvariable X eine stetige Verteilung hat, so gilt

$$F_X(\xi_\alpha) = \int\limits_{-\infty}^{\xi_\alpha} f_X(x)dx = \alpha$$

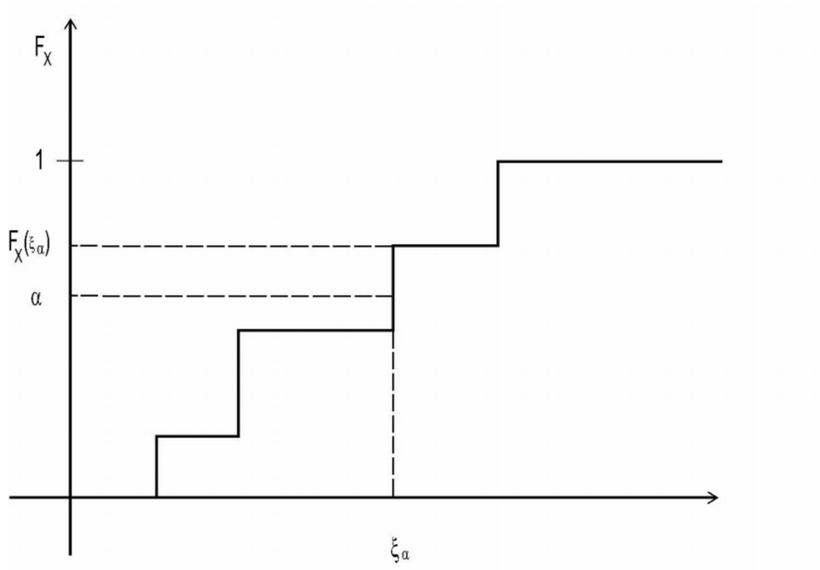
Wir wir bereits wissen, kann man das Integral als den Flächeninhalt unter einer Kurve interpretieren. In diesem Fall ist die Kurve die Wahrscheinlichkeitsfunktion. Das heisst $F_X(\xi_\alpha)$ ist der Fläche zwischen der x-Achse und der Wahrscheinlichkeitsfunktion f_x im Bereich von $-\infty$ bis ξ_α. Dieser Flächeninhalt ist gleich α.
Dazu eine Grafik:



- Ist die Verteilung von X diskret, so gilt

$$F_X(\xi_\alpha) \geq \alpha$$

Im folgenden Fall ist $F_X(\xi_\alpha)$ echt größer als α:



Streuungsparameter

Die **Varianz** einer Zufallsvariable X ist definiert als

$$VarX = E(X - EX)^2.$$

Anhand der Formel sieht man, dass die Varianz der Erwartungswert der quadratischen Abweichung einer Zufallsvariabe von ihrem Erwartungswert ist.

Häufig verwendet man für die Varianz auch das Symbol σ^2.

Die **Standardabweichung** ist die Wurzel aus der Varianz, also

$$\sigma = \sqrt{VarX}.$$

Die Varianz bzw. die Standardabweichung sind also Maße für die absolute Größe der Streuungen um den Erwartungswert.

Zur Berechnung der Varianzen sind folgende Regeln nützlich:

1. $Var(X) = E(X)^2 - (E(X))^2$
2. $Var(\alpha X + \beta) = \alpha^2 Var(X)$
3. Sind $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariablen, dann gilt

   $Var(X_1+X_2+\ldots+X_n) = Var(X_1)+Var(X_2)+\ldots+Var(X_n)$

Ein Beispiel: Sei X exponentialverteilt mit Parameter λ, dann gilt wegen $E(X) = \frac{1}{\lambda}$

$$\begin{aligned} E(X)^2 &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx = [-x^2 e^{-\lambda x}]_0^\infty + \underbrace{\int_0^\infty 2x e^{-\lambda x} dx}_{2*E(X)} \\ &= 0 + \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2} \end{aligned}$$

Das heisst

$$Var(X) = E(X)^2 - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Momente von Zufallsvariablen

Weitere die Verteilung einer Zufallsvariable charkterisierenden Elemente sind:

- **Momente**: $E(X)^k$
- **Zentrale Momente** $E(X - E(X))^k$

Das erste Moment $E(X)^1$ ist der Erwartungswert, das zweite zentrale Moment $E(X - E(X))^2$ kennen wir als Varianz. Aud den (zentralen) Momenten werden wichtige Grössen abgeleitet, nämlich

- **Schiefe**:

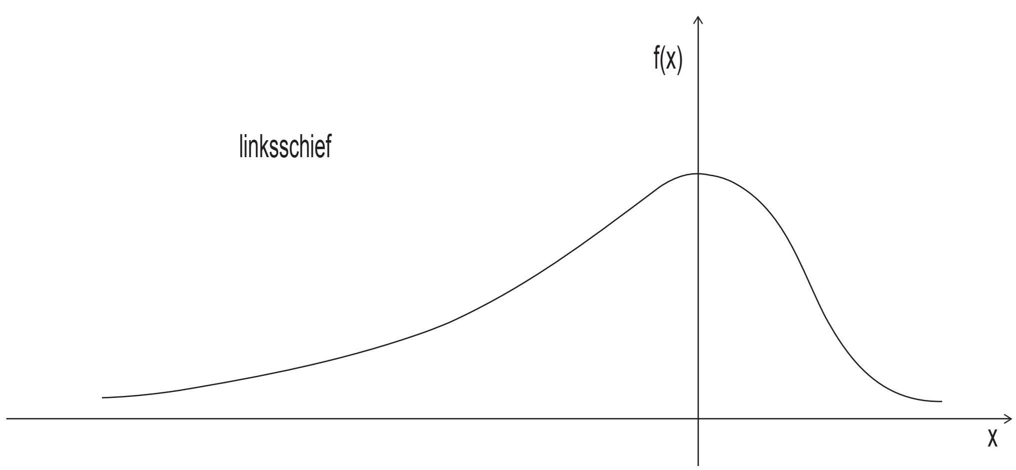
$$\frac{E(X-E(X))^3}{\sqrt{Var(X)}^3}$$

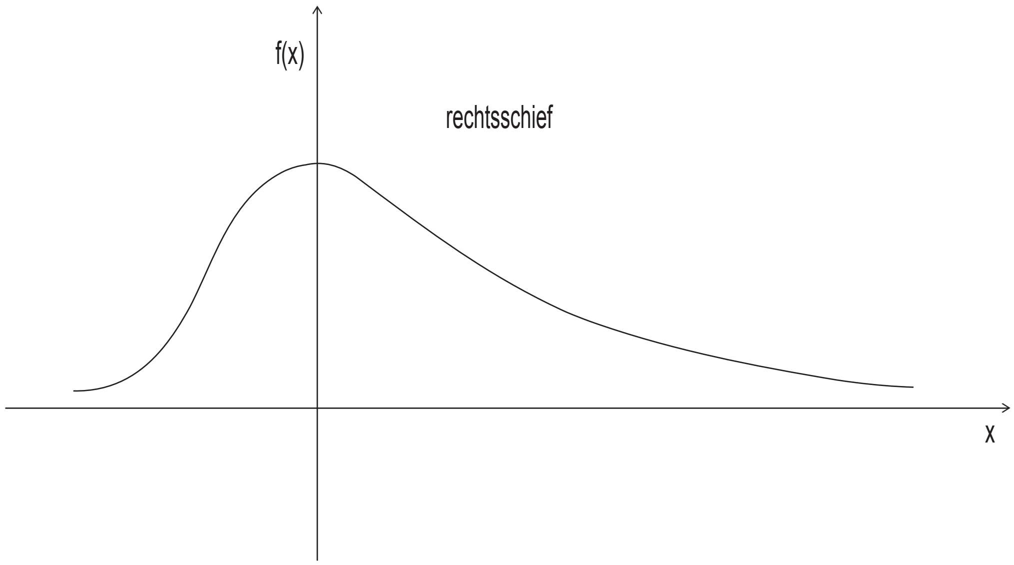
Eine Schiefe von 0 besitzen symmetrische Verteilungen (z.B. die Normalverteilung). Eine Verteilung die negative Schiefe besitzt wird "linksschief" genannt, eine mit positiver Schiefe "rechtsschief". Eine kleine Hilfe zur Identifizierung von links- bzw. rechtsschiefen Funktionen:

  - Eine linksschiefe Funktion wird auch **rechtssteil** genannt, ist sozusagen 'rechts steiler'
  - Eine rechtssschiefe Funktion wird auch **linkssteil** genannt, ist dementsprechend 'links steiler'

Beispiel:

Folgend sind jeweils eine linksschiefe (bzw. rechtssteile) und eine rechtsschiefe (bzw. linkssteile) Funktion:





- **Exzeß**
$$\frac{E(X - E(X))^4}{(Var(X))^2 - 3}$$

  Der Exzeß kann als Maß dafür angesehen werden, wie stark sich die Wölbung von der der Normalverteilung unterscheidet.

Kovarianz und Korrelation von Zufallsvariablen

Die **Kovarianz** ist eine Kenngröße, die den Zusammenhang zwischen zwei Zufallsvariablen (X und Y) misst. Sie ist definiert durch

$$Cov(X, Y) = E(X - E(X))(Y - E(Y)).$$

Es gelten folgende Rechenregeln:

- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $Cov(X, Y + Z) = Cov(X, Y) + Cov(X, Z)$

Zu beachten:

- Ist die Kovarianz gleich Null, so nennt man die Zufallsvariablen *unkorreliert*
- Zwei unabhängige Zufallsvariablen sind ebenfalls stets unkorreliert
- ACHTUNG: aus der Unkorreliertheit zweier Zufallsvariablen folgt NICHT ZWINGENDERMASSEN die Unabhängigkeit!

Die Korrelation ist eine *Normierung* der Kovarianz, welche folgendermaßen definiert ist:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

Aufgrund der Normierung liegt der Wert der Korrelation zweier Zufallsvariablen stets zwischen -1 und 1.