§65: ZUFALLSVARIABLE, ERWARTUNGSWERT, VARIANZ

65.1. Motivation

Oft möchte man dem Resultat eines Zufallsexperiments eine reelle Zahl zuordnen. Der Gewinn bei einem Glücksspiel ist ein Beispiel hierfür. In diesem Fall interessiert man sich auch für den zu erwartenden Gewinn und für ein Maß für die statistischen Schwankungen. Dies führt uns auf Begriffe wie Zufallsvariable, Erwartungswert und Varianz. In der Informatik werden sie u.a. bei der Zuverlässigkeitsanalyse von Systemen benötigt.

65.2. Def.:

Sei Ω ein Stichprobenraum. Eine Funktion X, die jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet heißt Zufallsvariable.

Bem.: Eine Zufallsvariable ist also weder zufällig noch eine Variable, sondern eine Funktion. Man kann sie stets als Gewinn bei einem Glücksspiel interpretieren.

65.3 Beispiel

Eine faire Münze mit Seiten 0 und 1 werde 3 Mal geworfen. Die Anzahl der Einsen sei der Gewinn.
Man kann $\Omega = \{000, 001, \ldots, 111\}$ als Stichprobenraum und den Gewinn als Zufallsvariable $X(\omega)$ auffassen.

Ergebnis ω	000	001	010	011	100	101	110	111
Gewinn $X(\omega)$	0	1	1	2	1	2	2	3
Ws. $P(\omega)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Versch. Ergebnisse ω_1, ω_2 können zum selben Gewinn führen.

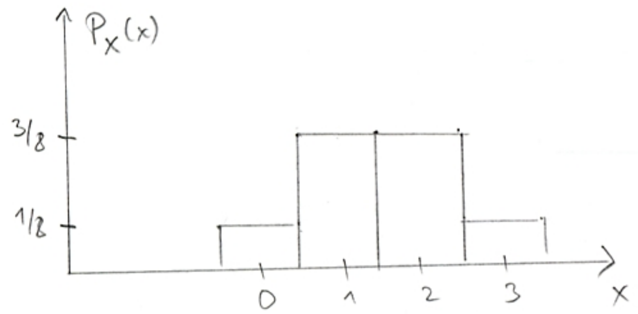
65.4. Def.: Die <u>Verteilung</u> P_X einer Zufallsvariablen X ordnet jedem Wert $x \in X$ eine Wahrscheinlichkeit $P_X(x)$ zu.

65.5. <u>Beispiel</u>

In Beispiel 65.3. können wir dem Gewinn folgende Wahrscheinlichkeiten machen.

Gewinn x	0	1	2	3
Ws. $P_X(x)$	$1/8$	$3/8$	$3/8$	$1/8$

Oft veranschaulicht man die Verteilung einer Zufallsvariablen durch ein <u>Histogramm</u>:



Bem.: Da wir hier meistens diskrete Zufallsvariablen betrachten, sind die Verteilungen ebenfalls diskret.

Interessiert man sich für den Durchschnittsgewinn je Versuchswiederholung, gelangt man zum Begriff des Erwartungswerts.

65.6. Def.: Unter dem <u>Erwartungswert</u> $E(X)$ einer (diskreten) Zufallsvariablen X versteht man das gewichtete Mittel der Funktion X über Ω, wobei jeder Wert mit seiner Wahrscheinlichkeit gewichtet wird:

$$\boxed{E(X) = \sum_{w \in \Omega} X(w)\, P(w)}$$

65.7. <u>Beispiel</u>

Für die Zufallsvariable X aus Beispiel 65.3. erhält man als Erwartungswert:

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8}$$

$$= \frac{12}{8} = \frac{3}{2}.$$

Bequemer hätte man den Erwartungswert mit Hilfe der Verteilung P_X berechnet (vgl. Tabelle in 65.5):

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Es gilt also auch

$$\boxed{E(X) = \sum_{x \in X(\Omega)} x\, P_X(x)}$$

Bem.: Bei kontinuierlichen Zufallsvariablen verwendet man Integrale statt Summen.

Mit dem Erwartungswert kann man gut arbeiten, denn es gilt:

65.8. Satz (Linearität des Erwartungswerts)

Seien X, Y Zufallsvariablen und $\alpha, \beta \in \mathbb{R}$. Dann gilt:

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$$

Beweis:
$$\begin{aligned} E(\alpha X + \beta Y) &= \sum_{\omega \in \Omega} (\alpha X(\omega) + \beta Y(\omega)) P(\omega) \\ &= \alpha \sum_{\omega \in \Omega} X(\omega) P(\omega) + \beta \sum_{\omega \in \Omega} Y(\omega) P(\omega) \\ &= \alpha E(X) + \beta E(Y) \end{aligned}$$
$\square$.

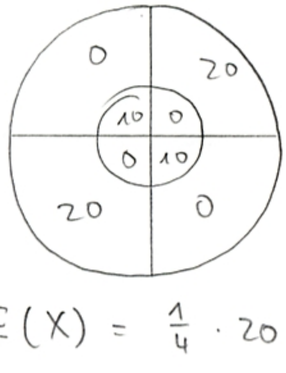
65.9. Unabhängigkeit zweier Zufallsvariablen

Mit 65.8 wissen wir, dass gilt: $E(X+Y) = E(X) + E(Y)$.
gilt jedoch auch $E(XY) = E(X)E(Y)$?
Man kann zeigen, dass dies dann gilt, wenn X und Y unabhängig sind, d.h. (vgl. 64.2):

$$P((X=a) \cap (Y=b)) = P(X=a) \cdot P(Y=b) \quad \forall a, b.$$

65.10. Beispiele

a)



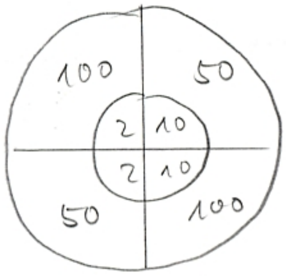
Glücksrad mit 4 Ergebnissen, auf denen die Zufallsvariable X (äußerer Ring) und Y (innerer Ring) definiert sind.

$$\left.\begin{aligned} E(X) &= \tfrac{1}{4} \cdot 20 + \tfrac{1}{4} \cdot 0 + \tfrac{1}{4} \cdot 20 + \tfrac{1}{4} \cdot 0 = 10 \\ E(Y) &= \tfrac{1}{4} \cdot 0 + \tfrac{1}{4} \cdot 10 + \tfrac{1}{4} \cdot 0 + \tfrac{1}{4} \cdot 10 = 5 \end{aligned}\right\} \quad E(X)E(Y) = 50$$

aber: $E(XY) = \frac{1}{4} \cdot 20 \cdot 0 + \frac{1}{4} \cdot 0 \cdot 10 + \frac{1}{4} \cdot 20 \cdot 0 + \frac{1}{4} \cdot 0 \cdot 10 = 0 \neq E(X)E(Y)$

X und Y sind nicht unabhängig:
Das Ereignis $Y=0$ hat die Ws. $\frac{1}{2}$. Weiß man jedoch, dass $X=20$ eingetreten ist, dann hat $Y=0$ die Ws. 1.

b)



$$E(X) = \frac{1}{4}\cdot 50 + \frac{1}{4}\cdot 100 + \frac{1}{4}\cdot 50 + \frac{1}{4}\cdot 100 = 75$$

$$E(Y) = \frac{1}{4}\cdot 10 + \frac{1}{4}\cdot 10 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 6$$

$$\Rightarrow \quad E(X)E(Y) = 450.$$

$$E(XY) = \frac{1}{4}\cdot 50\cdot 10 + \frac{1}{4}\cdot 100\cdot 10 + \frac{1}{4}\cdot 50\cdot 2 + \frac{1}{4}\cdot 100\cdot 2$$

$$= 450 = E(X)E(Y).$$

X und Y sind unabhängig:
$Y=2$ und $Y=10$ sind gleich wahrscheinlich. Weiß man, z.B., dass $X=50$ eingetreten ist, so sind $Y=2$ und $Y=10$ noch stets gleich wahrscheinlich. □

Oft ist man nicht nur am Erwartungswert interessiert. Man möchte auch wissen, wie stark die Verteilung um den Erwartungswert streut. Hierzu dienen die Begriffe Varianz und Standardabweichung:

65.11. Def.: Sei X eine Zufallsvariable mit Erwartungswert $\mu := E(X)$. Dann versteht man unter der Varianz $V(X) = \sigma^2$ den Erwartungswert von $(X-\mu)^2$:

$$\sigma^2 = V(X) = E((X-\mu)^2).$$

$\sigma = \sqrt{V(X)}$ nennt man die Standardabweichung (Streuung) von X.

65.12. Berechnung der Varianz

Wegen der Linearität des Erwartungswerts und wegen $E(\text{const.}) = \text{const.}$ gilt:

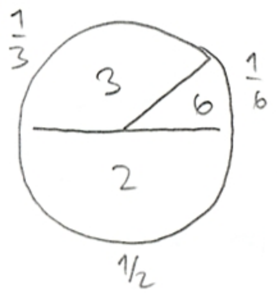
$$E((X-\mu)^2) = E(X^2 - 2\mu X + \mu^2)$$
$$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$
$$= E(X^2) - \mu^2$$

Wir erhalten somit eine wichtige Formel zur Berechnung der Varianz:

$$\boxed{\sigma^2 = E(X^2) - \mu^2} \qquad (\text{Verschiebungssatz})$$

65.13. Beispiel

Sei X der Gewinn auf dem Glücksrad



$$\Rightarrow \quad \mu = E(X) = \frac{1}{2}\cdot 2 + \frac{1}{3}\cdot 3 + \frac{1}{6}\cdot 6 = 3 \qquad \text{mittl. Gewinn}$$

$$E(x^2) = \frac{1}{2} 4 \cdot \frac{1}{3} 9 + \frac{1}{6}\cdot 36 = 11$$

$$\sigma^2 = E(X^2) - \mu^2 = 11 - 3^2 = 2 \qquad \text{Varianz}$$

$$\sigma = \sqrt{2} \qquad \text{Standardabweichung}$$

65.14. Satz (Eigenschaften der Varianz)

Sei X eine Zufallsvariable, und seien $\alpha, \beta \in \mathbb{R}$. Dann gilt:

i) $V(\alpha X) = \alpha^2 V(X)$

ii) $V(X+\beta) = V(X)$

Beweis:

i) $V(\alpha X) \overset{65.8}{=} E\left((\alpha X - \alpha\mu)^2\right) = E\left(\alpha^2 (X-\mu)^2\right) \overset{65.8}{=} \alpha^2 V(X)$

ii) $V(X+\beta) = E\left((X + \cancel{\beta} - \mu - \cancel{\beta})^2\right) = V(X)$ □.

65.15 → siehe S. 471

Wir wollen nun weitere wichtige Eigenschaften des Erwartungswerts studieren.

Aus dem Verschiebungssatz 65.12 folgt wegen $\sigma^2 \geq 0$, dass

$$E(X^2) - (E(X))^2 \geq 0$$

und somit $E(X^2) \geq (E(X))^2$. Dies ist ein Spezialfall eines allgemeineren Resultats:

65.16. Satz (Ungleichung von Jensen)

Sei $r: \mathbb{R} \to \mathbb{R}$ eine konvexe (!) Funktion und X eine Zufallsvariable. Dann gilt: $\boxed{E(r(X)) \geq r(E(X))}$

Beweis: Sei X zunächst eine diskrete Zufallsvariable.

Dann fassen wir $E(X) = \sum_{w \in \Omega} X(w) P(w)$ als Konvexkombination der $X(w)$, $w \in \Omega$ mit Gewichten $P(w)$ auf (d.h. $P(w) \geq 0$ $\forall w \in \Omega$ und $\sum_{w \in \Omega} P(w) = 1$).

Aus der Konvexität von r folgt dann mit 25.5:

$$E(r(X)) = \sum_{\omega\in\Omega} r(X(\omega))\, P(\omega)$$
$$\geq r\Big(\sum_{\omega\in\Omega} X(\omega)\, P(\omega)\Big) = r(E(X))$$

Ist X eine kontinuierliche Zufallsvariable, ersetzt man Summen durch Integrale. □

65.17. <u>Beispiele</u>

a) Die Funktion $r(t) = t^2$ ist konvex, da $r''(t) = 2 > 0$.
Daher gilt: $$E(X^2) \geq (E(X))^2.$$

b) Sei $\theta > 0$. Dann ist $r(t) = e^{\theta t}$ konvex und es gilt
$$E(e^{\theta X}) \geq e^{\theta E(X)}.$$
Die Funktion $E(e^{\theta X})$ wird später eine wichtige Rolle spielen.

65.18. <u>Gleichung von Wald</u>

Seien $X_1, X_2, \ldots$ unabhängige Zufallsvariablen.
Manchmal interessiert man sich für die erste Zeit (<u>Stoppzeit</u>), $N = n$, in der die Summe $X_1 + \ldots + X_n$ einen vorgegebenen Wert y übersteigt, d.h. $\sum_{i=1}^{n-1} X_i \leq y$ und $\sum_{i=1}^{n} = y$. Dann ist N selbst wieder eine Zufallsvariable. Für ihren Erwartungswert kann man zeigen:

Satz (<u>Gleichung von Wald</u>)

Seien $X_1, X_2, \ldots$ unabhängige, <u>gleich verteilte</u> Zufallsvariablen (d.h. $P(\omega)$ ist gleich $\forall\, \omega \in \Omega$) mit endlichem Erwartungswert, und sei N eine Stoppzeit für $X_1, X_2, \ldots$. Ferner sei $S_N = \sum_{i=1}^{N} X_i$.

Dann gilt: $$E(S_N) = E(X)\, E(N).$$

65.19. Beispiel

Sei $X_i = 1$ falls beim i-ten Münzwurf Kopf eintritt, ansonsten 0.
Wie ist der Erwartungswert $E(N)$ für die Stoppzeit

$$N := \min\{n \mid X_1 + \ldots + X_N \geq 10\} \quad ?$$

Es gilt $E(X) = \frac{1}{2}$ und $E(S_N) = 10$.

$\Rightarrow E(N) = 20$.

65.15. Standardisierte Zufallsvariablen

Ist X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2, so nennt man

$$X^* := \frac{X-\mu}{\sigma}$$

die Standardisierte von X. Es gilt:

$$E(X^*) = \frac{1}{\sigma} E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$V(X^*) \overset{65.14}{=} \frac{1}{\sigma^2} V(X) = \frac{\sigma^2}{\sigma^2} = 1.$$

Eine solche Standardisierung ist nützlich, wenn man die Verteilung einer Zufallsvariablen mit einer tabellierten Verteilung vergleichen möchte, da letztere oft nur in standardisierter Form vorliegt.

Eng verwandt mit dem Begriff der Varianz ist die Kovarianz:
Sie ist ein Maß für die Unabh. zweier Zufallsvariablen.

65.20. Def.: Seien X, Y Zufallsvariablen mit Erwartungswert μ_X, μ_Y und Varianz $\sigma_X^2, \sigma_Y^2 > 0$. Dann heißt

$$\mathrm{Cov}(X,Y) := \sigma_{XY}^2 := E\big((X-\mu_X)(Y-\mu_Y)\big)$$

die Kovarianz von X und Y, und

$$\varrho_{XY} := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

den <u>Korrelationskoeffizienten</u> von X und Y.
Ist $Cov(X,Y) = 0$, so heißen X und Y <u>unkorreliert</u>.

65.21. <u>Bemerkungen</u>

a) $V(X) = Cov(X,X)$

b) Sind $X-\mu_X$ und $Y-\mu_Y$ Funktionen auf einem endlichen Stichprobenraum $\Omega = \{w_1, \dots, w_n\}$, so kann man sie auch als Vektoren im $\mathbb{R}^n$ mit der i-ten Komponente $X(w_i)-\mu_X$ bzw. $Y(w_i)-\mu_Y$ interpretieren.
Dann bezeichnen (vgl. §41, §43):
- die Standardabweichungen die euklidischen Normen
- die Varianzen die quadrierten eukl. Normen
- die Kovarianz das euklidische Skalarprodukt
- der Korrelationskoeffizient den Arcuscosinus des Winkels zwischen beiden Vektoren.
  Insbes. ist $-1 \le \varrho_{XY} \le 1$.
- die Unkorreliertheit die Orthogonalität

wenn wir das gewichtete euklidische Skalarprodukt

$$\langle u,v \rangle := \sum_{i=1}^{n} p_i u_i v_i$$

zu Grunde legen. Dabei ist $p_i = P(w_i)$.

Man kann Folgendes zeigen:

65.22. Satz (<u>Rechenregeln für die Kovarianz</u>)

Seien X, Y Zufallsvariablen mit Erwartungswert μ_X, μ_Y. Dann gilt:

a) $Cov(X,Y) = E(X,Y) - \mu_X \mu_Y$ (vgl. 65.12)

b) $Cov(\alpha X + \beta, \gamma Y + \delta) = \alpha\gamma\, Cov(X,Y)$ (vgl. 65.14).

c) $Cov(X,Y) = Cov(Y,X)$

d) Für n Zufallsvariable $X_1, \ldots, X_n$ gilt:

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$$

e) Sind X, Y unabhängig, so sind sie auch unkorreliert.

f) Für <u>unabhängige</u> $X_1, \ldots, X_n$ gilt:

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i)$$

65.23. <u>Bemerkungen</u>

a) Die Umkehrung von 65.23 (e) gilt nicht! Beispiel:

Ergebnis w	1	2	3	4
Zufallsvar. X	1	−1	2	−2
" Y	−1	1	2	−2
Wahrscheinl.	2/5	2/5	1/10	1/10

Dann ist $E(X) = 0 = E(Y)$ und

$Cov(X, Y) = -1 \cdot \frac{2}{5} - 1 \cdot \frac{2}{5} + 4 \cdot \frac{1}{10} + 4 \cdot \frac{1}{10} = 0$, aber:

X und Y sind nicht unabhängig, denn $X(w)$ bestimmt w und $Y(w)$ eindeutig.

b) In der Informatik tauchen Erwartungswerte und Varianzen z.B. bei der Zuverlässigkeitsanalyse von Systemen auf oder bei der Abschätzung von Wartezeiten bei Internetanfragen.
Kovarianzen sind u.A. wichtig im Bereich des maschinellen Lernens.

65.24. Zuverlässigkeitsanalyse von Systemen

Ein System bestehe aus n Komponenten. Der Zustand der k-ten Komponente wird durch die Zufallsvariable (Indikator)

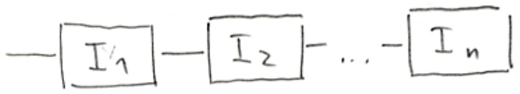
$$I_k = \begin{cases} 1 & (k\text{-te Komp. funktioniert}) \\ 0 & (\text{" " " nicht}) \end{cases}$$

beschrieben. Ihr Erwartungswert beschreibt die Zuverlässigkeit der Komponente k. Wir setzen:

$$p_k := E(I_k), \qquad q_k := 1 - p_k \quad (\text{Ausfallws.})$$

Interessiert man sich für die Zuverlässigkeit p des Gesamtsystems, muss man verschiedene Fälle unterscheiden:

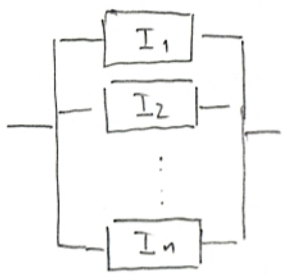
a) Reihensysteme



Ein Reihensystem arbeitet, wenn alle Komponenten arbeiten:

$$p = p_1 \cdot \ldots \cdot p_n$$

b) Parallelsysteme



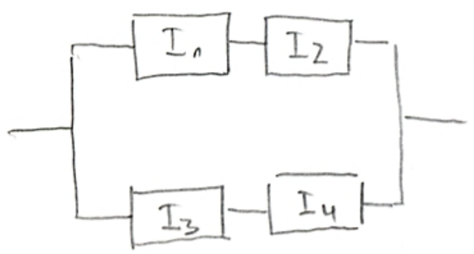
Ein Parallelsystem fällt aus, wenn alle Komponenten ausfallen:

$$q = q_1 \cdot \ldots \cdot q_n$$

$$\Longrightarrow \quad p = 1 - q = 1 - (1-p_1) \cdot \ldots \cdot (1-p_n)$$

c) Gemischte Systeme

werden hierarchisch in Reihensyst. u. Parallelsysteme zerlegt.



oberes Reihensystem: $p_1 p_2$

unteres " : $p_3 p_4$

äußeres Parallelsystem: $p = 1 - (1 - p_1 p_2)(1 - p_3 p_4)$.