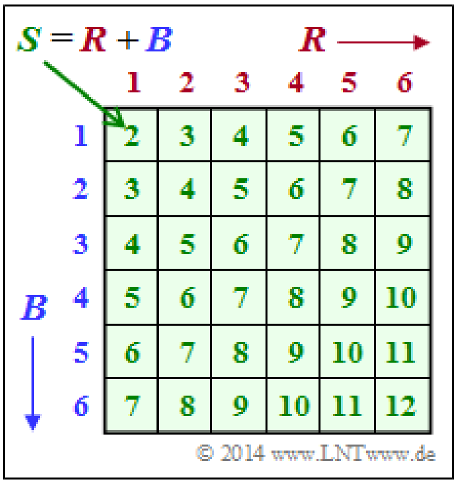
A3.1: Wahrscheinlichkeiten beim Würfeln



Wir betrachten das Zufallsexperiment „Würfeln mit ein oder zwei Würfeln”. Beide Würfel sind fair (die sechs möglichen Ergebnisse sind gleichwahrscheinlich) und durch ihre Farben unterscheidbar:

- Die Zufallsgröße $R = \{1, 2, 3, 4, 5, 6\}$ bezeichnet die Augenzahl des roten Würfels.
- Die Zufallsgröße $B = \{1, 2, 3, 4, 5, 6\}$ bezeichnet die Augenzahl des blauen Würfels.
- Die Zufallsgröße $S = R + B$ steht für die Summe beider Würfel.

In dieser Aufgabe sollen verschiedene Wahrscheinlichkeiten mit Bezug zu den Zufallsgrößen R, B und S berechnet werden, wobei das oben angegebene Schema hilfreich sein kann. Dieses beinhaltet die Summe S in Abhängigkeit von R und B.

Hinweis: Die Aufgabe dient zur Vorbereitung für weitere Aufgaben zum **Kapitel 3.1** dieses Buches „Einführung in die Informationstheorie”. Wiederholt wird hier insbesondere der Lehrstoff von **Kapitel 1.1** und **Kapitel 1.3** des Buches „Stochastische Signaltheorie”.

Fragebogen zu "A3.1: Wahrscheinlichkeiten beim Würfeln"

a) Geben Sie die folgenden Wahrscheinlichkeiten an:

Pr(R = 6) =

Pr($B \le 2$) =

Pr(R = B) =

b) Wie lauten die folgenden Wahrscheinlichkeiten?

Pr(S = 3) =

Pr(S = 7) =

Pr(S ist ungeradzahlig) =

c) Geben Sie die folgenden Wahrscheinlichkeiten an:

Pr[(R = 6) ∪ (B = 6)] =

Pr[(R = 6) ∩ (B = 6)] =

d) Wie groß ist die Wahrscheinlichkeit, dass beim L–ten Doppelwurf zum ersten Mal eine „6” dabei ist?

Pr(erste „6” beim 1. Wurf) = (L=1)

Pr(erste „6” beim 2. Wurf) = (L=2)

Pr(erste „6” beim 3. Wurf) = (L=3)

e) Wie groß ist die Wahrscheinlichkeit „Um die erste „6” zu erhalten, benötigt man eine geradzahlige Anzahl an Doppelwürfen”? Mit der Nomenklatur gemäß (d):

Pr(L ist geradzahlig) =

Z3.1: Karten ziehen

Aus einem Kartenspiel mit 32 Karten, darunter 4 Asse, werden nacheinander 3 Karten herausgezogen. Für Frage (a) wird vorausgesetzt, dass nach dem Ziehen einer Karte diese in den Stapel zurückgelegt wird, dieser neu gemischt wird und anschließend die nächste Karte gezogen wird.

Dagegen sollen Sie für die weiteren Teilfragen ab (b) davon ausgehen, dass die drei Karten auf einmal gezogen werden („Ziehen ohne Zurücklegen“).

Im Folgenden bezeichnen wir mit A_i das Ereignis, dass die zum Zeitpunkt i gezogene Karte ein Ass ist. Hierbei ist i = 1, 2, 3 zu setzen. Das Komplementärereignis sagt dann aus, dass zum Zeitpunkt i irgend eine andere Karte gezogen wird.

Hinweis: Die Aufgabe behandelt den Lehrstoff von **Kapitel 1.3** im Buch „Stochastische Signaltheorie”. Sie wird hier zur Vorbereitung auf die ähnliche Thematik von **Kapitel 3.1** des Buches „Einführung in die Informationstheorie” wiederholt.

Eine Zusammenfassung der theoretischen Grundlagen mit Beispielen bringt folgendes Lernvideo:

Statistische (Un-)Abhängigkeit (3-teilig: Dauer Teil 1: 4:20 – Teil 2: 3:40 – Teil 3: 3:40)

Fragebogen zu "Z3.1: Karten ziehen"

a) Betrachten Sie zunächst den Fall „Ziehen mit Zurücklegen“. Wie groß ist die Wahrscheinlichkeit p_a, dass drei Asse gezogen werden?

$\boldsymbol{p_\text{a}}$ =

b) Mit welcher Wahrscheinlichkeit p_b werden drei Asse gezogen, wenn man die Karten nicht zurücklegt? Warum ist p_b kleiner/gleich/größer als p_a?

$\boldsymbol{p_\text{b}}$ =

c) Betrachten Sie weiterhin den Fall „Ziehen ohne Zurücklegen“. Wie groß ist die Wahrscheinlichkeit p_c, dass kein einziges Ass gezogen wird?

$\boldsymbol{p_\text{c}}$ =

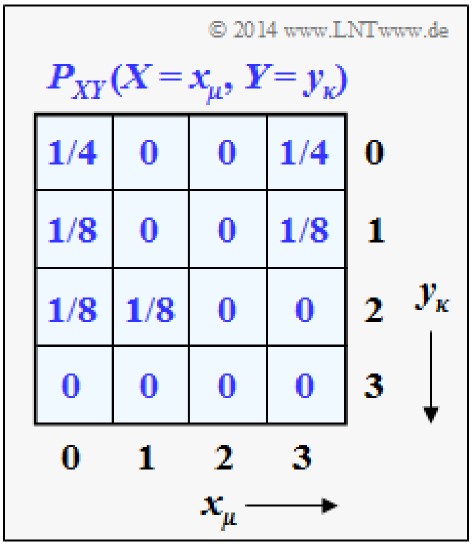
d) Wie groß ist die Wahrscheinlichkeit p_d, dass genau ein Ass gezogen wird?

$\boldsymbol{p_\text{d}}$ =

e) Wie groß ist die Wahrscheinlichkeit, dass zwei der drei gezogenen Karten Asse sind? *Hinweis*: Die Ereignisse „genau i Asse werden gezogen” mit i = 0, 1, 2, 3 beschreiben ein vollständiges System.

$\boldsymbol{p_\text{e}}$ =

A3.2: Erwartungswertberechnungen



Wir betrachten folgende Wahrscheinlichkeitsfunktionen:

- $P_X(X) = [1/2,\ 1/8,\ 0,\ 3/8]$,
- $P_Y(Y) = [1/2,\ 1/4,\ 1/4,\ 0]$,
- $P_U(U) = [1/2,\ 1/2]$,
- $P_V(V) = [3/4,\ 1/4]$.

Für die dazugehörigen Zufallsgrößen gelte:

- $X = \{0, 1, 2, 3\}$, $Y = \{0, 1, 2, 3\}$,
- $U = \{0, 1\}$, $V = \{0, 1\}$.

Oft muss man in der Informationstheorie für solche diskreten Zufallsgrößen verschiedene Erwartungswerte der Form

$$\mathrm{E}\left[F(X)\right] = \sum_{x \,\in\, \mathrm{supp}(P_X)} P_X(x) \cdot F(x)$$

berechnen. Hierbei bedeuten:

- $P_X(X)$ bezeichnet die *Wahrscheinlichkeitsfunktion* der diskreten Zufallsgröße X.
- Der *Support* von P_X umfasst alle diejenigen Realisierungen x der Zufallsgröße X mit nicht verschwindender Wahrscheinlichkeit. Formal kann hierfür geschrieben werden:

  $$\mathrm{supp}(P_X) = \{x:\ \ x \in X \ \underline{\mathrm{und}}\ P_X(x) \ne 0\}\,.$$

- $F(X)$ ist eine (beliebige) reellwertige Funktion, die im gesamten Definitionsgebiet der Zufallsgröße angebbar ist.

In der Aufgabe sollen die Erwartungswerte für verschiedene Funktionen $F(X)$ berechnet werden, unter anderem für

- $F(X) = 1/P_X(X)$,
- $F(X) = P_X(X)$,
- $F(X) = -\log_2 P_X(X)$.

Hinweis: Die Aufgabe gehört zum Themengebiet von **Kapitel 3.1** dieses Buches. Weiter gilt, ohne dass dies für die Lösung dieser Aufgabe von Interesse ist:

- Die beiden „eindimensionalen” Wahrscheinlichkeitsfunktionen $P_X(X)$ und $P_Y(Y)$ ergeben sich aus der dargestellten 2D–PMF $P_{XY}(X, Y)$, wie in **Aufgabe Z3.2** gezeigt werden soll.
- Zu den binären Wahrscheinlichkeitsfunktionen $P_U(U)$ und $P_V(V)$ kommt man entsprechend den Modulo–Operationen $U = X \bmod 2$ sowie $V = Y \bmod 2$ (siehe ebenfalls Aufgabe Z3.2)

Fragebogen zu "A3.2: Erwartungswertberechnungen"

a) Welche Ergebnisse liefern die folgenden Erwartungswerte?

$\mathbf{E}[1/P_X(X)]$ =

$\mathbf{E}[1/P_Y(Y)]$ =

b) Geben Sie die folgenden Erwartungswerte an:

$\mathbf{E}[P_X(X)]$ =

$\mathbf{E}[P_Y(Y)]$ =

c) Berechnen Sie nun die folgenden Erwartungswerte:

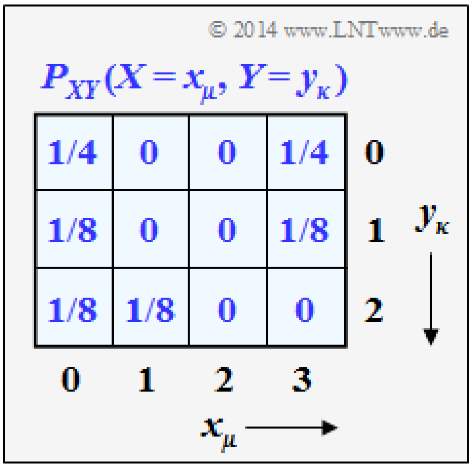
$\mathbf{E}[P_Y(X)]$ =

$\mathbf{E}[P_X(Y)]$ =

d) Welche der folgenden Aussagen sind zutreffend?

- [ ] $\mathrm{E}[-\log_2 P_U(U)]$ ergibt die Entropie der Zufallsgröße U.
- [ ] $\mathrm{E}[-\log_2 P_V(V)]$ ergibt die Entropie der Zufallsgröße V.
- [ ] $\mathrm{E}[-\log_2 P_V(U)]$ ergibt die Entropie der Zufallsgröße V.

Z3.2: 2D–Wahrscheinlichkeitsfunktion



Wir betrachten die Zufallsgrößen

$X = \{0, 1, 2, 3\},$

$Y = \{0, 1, 2\}.$

deren gemeinsame Wahrscheinlichkeitsfunktion $P_{XY}(X, Y)$ gegeben ist. Aus dieser 2D–Wahrscheinlichkeitsfunktion sollen die eindimensionalen Wahrscheinlichkeitsfunktionen $P_X(X)$ und $P_Y(Y)$ ermittelt werden. Man nennt eine solche manchmal auch Randwahrscheinlichkeit (englisch: *Marginal Probability*).

Gilt $P_{XY}(X, Y) = P_X(X) \cdot P_Y(Y)$, so sind die beiden Zufallsgrößen X und Y statistisch unabhängig. Andernfalls bestehen statistische Bindungen zwischen X und Y.

Im zweiten Teil der Aufgabe betrachten wir die Zufallsgrößen

$U = \{0, 1\}, \quad V = \{0, 1\},$

die sich aus X und Y durch Modulo–2–Operationen ergeben:

$U = X \bmod 2, \quad V = Y \bmod 2.$

Hinweis: Die Aufgabe gehört zum Themengebiet von **Kapitel 3.1.** Ausgegangen wird hier von der gleichen Konstellation wie in **Aufgabe A3.2.** Dort wurde die Zufallsgrößen $Y = \{0, 1, 2, 3\}$ betrachtet, allerdings mit dem Zusatz $\Pr(Y = 3) = 0$. Die so erzwungene Eigenschaft $|X| = |Y|$ war in Aufgabe A3.2 zur formalen Berechnung des Erwartungswertes $\mathrm{E}[P_X(Y)]$ von Vorteil.

Fragebogen zu "Z3.2: 2D–Wahrscheinlichkeitsfunktion"

a) Wie lautet die Wahrscheinlichkeitsfunktion $P_X(X)$?

$\boldsymbol{P_X(0)}$ =

$\boldsymbol{P_X(1)}$ =

$\boldsymbol{P_X(2)}$ =

$\boldsymbol{P_X(3)}$ =

b) Wie lautet die Wahrscheinlichkeitsfunktion $P_Y(Y)$?

$\boldsymbol{P_Y(0)}$ =

$\boldsymbol{P_Y(1)}$ =

$\boldsymbol{P_Y(2)}$ =

c) Sind die Zufallsgrößen X und Y statistisch unabhängig?

- Ja.
- Nein.

d) Ermitteln Sie die Wahrscheinlichkeiten $P_{UV}(U, V)$.

$\boldsymbol{P_{UV}(U = 0, V = 0)}$ =

$\boldsymbol{P_{UV}(U = 0, V = 1)}$ =

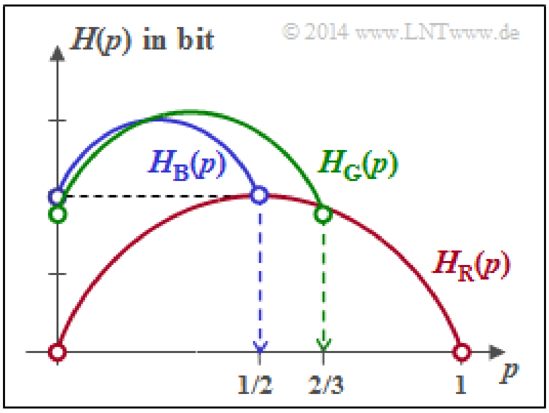
$\boldsymbol{P_{UV}(U = 1, V = 0)}$ =

$\boldsymbol{P_{UV}(U = 1, V = 1)}$ =

e) Sind die Zufallsgrößen U und V statistisch unabhängig?

- Ja.
- Nein.

A3.3: Entropie von Ternärgrößen



Rechts sehen Sie die Entropiefunktionen $H_R(p)$, $H_B(p)$ und $H_G(p)$, wobei „R" für „Rot" steht, usw. Für alle Zufallsgrößen lautet die Wahrscheinlichkeitsfunktion:

$$P_X(X) = [p_1\,,p_2\,,p_3] \quad\Rightarrow\quad |X| = 3\,.$$

Es gilt der Zusammenhang $p_1 = p$ und $p_2 = 1 - p_3 - p$.

Die Wahrscheinlichkeitsfunktion einer Zufallsgröße

$$X = \{\,x_1\,,\,x_2\,,\,\ldots\,,\,x_\mu\,,\,\ldots\,,\,x_M\,\}$$

mit dem Symbolumfang $|X| = M$ lautet allgemein:

$$P_X(X) = [\,p_1\,,\,p_2\,,\,\ldots\,,\,p_\mu\,,\,\ldots\,,\,p_M\,]\,.$$

Die Entropie (Unsicherheit) dieser Zufallsgröße berechnet sich entsprechend der Gleichung

$$H(X) = {\rm E}\,[\log_2 1/P_X(X)]\,,$$

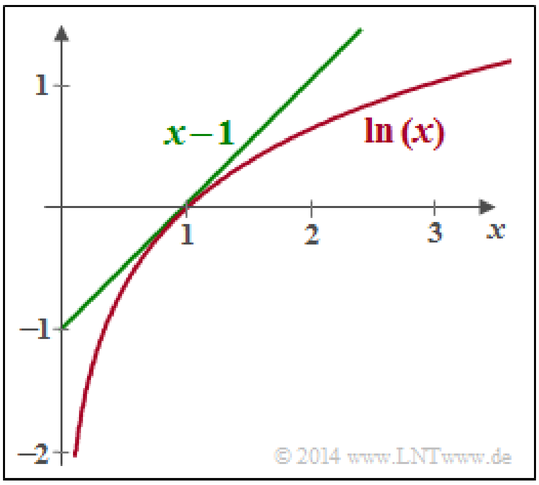
und liegt stets im Bereich $0 \le H(X) \le \log_2 |X|$. Die untere Schranke $H(X) = 0$ ergibt sich, wenn eine beliebige Wahrscheinlichkeit $p_\mu = 1$ ist und alle anderen 0 sind. Die obere Schranke soll hier wie in **[Kra13]** hergeleitet werden:

- Durch Erweiterung obiger Gleichung um $|X|$ in Zähler und Nenner erhält man unter Verwendung von $\log_2(x) = \ln(x)/\ln(2)$:

$$H(X) = \frac{1}{\ln(2)}\cdot {\rm E}\left[\ln \frac{1}{|X|\cdot P_X(X)}\right] + \log_2 |X|\,.$$

- Wie aus nachfolgender Grafik hervorgeht, gilt die Abschätzung $\ln(x) \le x - 1$ mit der Identität für $x = 1$. Somit kann geschrieben werden:

$$H(X) \le \frac{1}{\ln(2)}\cdot {\rm E}\left[\frac{1}{|X|\cdot P_X(X)} - 1\right] + \log_2 |X|\,.$$



- In **Aufgabe A3.2** wurde für den Fall, dass $p_\mu \ne 0$ für alle μ gilt, der Erwartungswert ${\rm E}[1/P_X(X)]$ zu $|X|$ berechnet. Damit verschwindet der erste Term und man erhält das bekannte Ergebnis:

$$H(X) \le \log_2 |X|\,.$$

Hinweis: Die Aufgabe gehört zu **Kapitel 3.1.** Es wird auf die binäre Entropiefunktion Bezug genommen:

$$H_{\rm bin}(p) = p\cdot \log_2 \frac{1}{p} + (1-p)\cdot \log_2 \frac{1}{1-p}\,.$$

Fragebogen zu "A3.3: Entropie von Ternärgrößen"

a) Welche Aussagen gelten für die rote Entropiefunktion $H_R(p)$?

- ☐ $H_R(p)$ ergibt sich z.B. mit $p_3 = 0$, $p_1 = p$, $p_2 = 1 - p$.
- ☐ $H_R(p)$ ist identisch mit der binären Entropiefunktion $H_{bin}(p)$.

b) Welche Eigenschaften weist die binäre Entropiefunktion auf?

- ☐ $H_{bin}(p)$ ist konkav hinsichtlich des Parameters p.
- ☐ Es gilt $\text{Max}[H_{bin}(p)] = 2$ bit.

c) Welche Aussagen gelten für die blaue Entropiefunktion $H_B(p)$?

- ☐ $H_B(p)$ ergibt sich z.B. mit $p_3 = 1/2$, $p_1 = p$, $p_2 = 1/2 - p$.
- ☐ Es gilt $H_B(p = 0) = 1$ bit.
- ☐ Es gilt $\text{Max}[H_B(p)] = \log_2 (3)$ bit.

d) Welche Aussagen gelten für die grüne Entropiefunktion $H_G(p)$?

- ☐ $H_G(p)$ ergibt sich z.B. mit $p_3 = 1/3$, $p_1 = p$, $p_2 = 2/3 - p$.
- ☐ Es gilt $H_G(p = 0) = 1$ bit.
- ☐ Es gilt $\text{Max}[H_G(p)] = \log_2 (3)$ bit.

Z3.3: $H(X)$ für verschiedene $P_X(X)$

$P_X(1)$	$P_X(2)$	$P_X(3)$	$P_X(4)$	Entropie
0.1	0.2	0.3	0.4	$H_a(X)$
0.1	0.2	p_3	p_4	$H_b(X)$
0.1	p_2	p_3	0.4	$H_c(X)$
p_1	p_2	p_3	p_4	$H_{max}(X)$

© 2014 www.LNTwww.de

In der ersten Zeile der nebenstehenden Tabelle ist die mit „a" bezeichnete Wahrscheinlichkeitsfunktion angegeben. Für dieses $P_X(X)$ soll in der Teilaufgabe (a) die Entropie

$$H_a(X) = \mathrm{E}\left[\log_2 \frac{1}{P_X(X)}\right]$$

berechnet werden. Da hier der Logarithmus zur Basis 2 verwendet wird, ist die Pseudo–Einheit „bit" anzufügen.

In den weiteren Aufgaben sollen jeweils einige Wahrscheinlichkeiten variiert werden und zwar derart, dass sich jeweils die größtmögliche Entropie ergibt:

- Durch geeignete Variation von p_3 und p_4 kommt man zur maximalen Entropie $H_b(X)$ unter der Voraussetzung $p_1 = 0.1$ und $p_2 = 0.2$ ⇒ Teilaufgabe (b).
- Durch geeignete Variation von p_2 und p_3 kommt man zur maximalen Entropie $H_c(X)$ unter der Voraussetzung $p_1 = 0.1$ und $p_4 = 0.4$ ⇒ Teilaufgabe (c).
- In der Teilaufgabe (d) sind alle vier Parameter zur Variation freigegeben, die entsprechend der maximalen Entropie ⇒ $H_{max}(X)$ zu bestimmen sind.

Hinweis: Die Aufgabe bezieht sich auf das **Kapitel 3.1.**

Fragebogen zu "Z3.3: $H(X)$ für verschiedene $P_X(X)$ "

a) Zu welcher Entropie führt $P_X(X)$ = [0.1, 0.2, 0.3, 0.4]?

$H_\mathrm{a}(X)$ = bit

b) Es gelte nun allgemein $P_X(X)$ = [0.1, 0.2, p_3, p_4]. Welche Entropie erhält man, wenn p_3 und p_4 bestmöglich gewählt werden?

$H_\mathrm{b}(X)$ = bit

c) Nun gelte $P_X(X)$ = [0.1, p_2, p_3, 0.4]. Welche Entropie erhält man, wenn p_2 und p_3 bestmöglich gewählt werden?

$H_\mathrm{c}(X)$ = bit

d) Welche Entropie erhält man, wenn alle Wahrscheinlichkeiten (p_1, p_2, p_3 und p_4) bestmöglich gewählt werden können?

$H_\mathrm{max}(X)$ = bit

A3.4: KLD & Binomialverteilung

Wir gehen hier von der **Binomialverteilung** aus, die durch die Parameter I und p gekennzeichnet ist ⇒ siehe Buch „Stochastische Signaltheorie":

- Wertebereich:

$$X = \{\,0,\,1,\,2,\,\ldots\,,\,\mu,\,\ldots\,,\,I\,\}\,,$$

- Wahrscheinlichkeiten:

$$P_X(X=\mu) = \binom{I}{\mu} \cdot p^{\mu} \cdot (1-p)^{I-\mu}\,,$$

- linearer Mittelwert:

$$m_X = I \cdot p\,,$$

- Varianz:

$$\sigma_X^2 = I \cdot p \cdot (1-p)\,.$$

μ	$P_X(X=\mu)$: Gegebene Binomialverteilung	$P_Y(Y=\mu)$: Poisson-Näherung mit				
		$\lambda = 0.5$	$\lambda = 0.8$	$\lambda = 1.0$	$\lambda = 1.2$	$\lambda = 2.0$
0	0.3277	0.6065	0.4493	0.3679	0.3012	0.1353
1	0.4096	0.3033	0.3595	0.3679	0.3614	0.2707
2	0.2048	0.0758	0.1438	0.1839	0.2169	0.2707
3	0.0512	0.0126	0.0383	0.0613	0.0867	0.1804
4	0.0064	0.0016	0.0077	0.0153	0.0260	0.0902
5	0.0003	0.0002	0.0012	0.0031	0.0062	0.0361
6	0	$<10^{-4}$	0.0002	0.0005	0.0012	0.0120
7	0	$<10^{-4}$	$<10^{-4}$	0.0001	0.0002	0.0034
8	0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0009
9	0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	0.0002
10	0	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

© 2014 www.LNTwww.de

Im rot hinterlegten Teil obiger Tabelle sind die Wahrscheinlichkeiten $P_X(X=\mu)$ der hier betrachteten Binomialverteilung angegeben. In der Teilaufgabe (a) sollen Sie die dazugehörigen Verteilungsparameter I und p bestimmen.

Diese vorgegebene Binomialverteilung soll hier durch eine **Poissonverteilung** Y approximiert werden, gekennzeichnet durch die Rate λ:

- Wertebereich:

$$Y = \{\,0,\,1,\,2,\,\ldots\,,\,\mu,\,\ldots\,\}\,,$$

- Wahrscheinlichkeiten:

$$P_Y(Y=\mu) = \frac{\lambda^{\mu}}{\mu!} \cdot \mathrm{e}^{\lambda}\,,$$

- Erwartungswerte:

$$m_Y = \sigma_Y^2 = \lambda\,.$$

Um abschätzen zu können, ob die Wahrscheinlichkeitsfunktion $P_X(X)$ ausreichend gut durch $P_Y(Y)$ approximiert wird, kann man auf die so genannten *Kullback–Leibler–Distanzen* (KLD) zurückgreifen, teilweise in der Literatur auch *relative Entropien* genannt. Angepasst an das vorliegende Beispiel lauten diese:

$$D(P_X\,||\,P_Y) \;=\; \mathrm{E}\left[\log_2 \frac{P_X(X)}{P_Y(X)}\right] \;=\; \sum_{\mu=0}^{I} P_X(\mu) \cdot \log_2 \frac{P_X(\mu)}{P_Y(\mu)}\,,$$

$$D(P_Y\,||\,P_X) \;=\; \mathrm{E}\left[\log_2 \frac{P_Y(X)}{P_X(X)}\right] \;=\; \sum_{\mu=0}^{\infty} P_Y(\mu) \cdot \log_2 \frac{P_Y(\mu)}{P_X(\mu)}\,.$$

Bei Verwendung des *Logarithmus dualis* (zur Basis 2) ist hierbei dem Zahlenwert die Pseudo–Einheit

„bit” hinzuzufügen.

In nebenstehender Ergebnistabelle ist die sog. Kullback–Leibler–Distanz $D(P_X||P_Y)$ in „bit” zwischen der Binomial–PMF P_X und einigen Poisson–Näherungen P_Y (mit fünf verschiedenen Raten λ) eingetragen. Die jeweilige Entropie $H(Y)$, die ebenfalls von der Rate λ abhängt, ist in der ersten Zeile angegeben.

Poisson-Näherung Y mit Rate λ ($m_Y = \sigma_Y^2 = \lambda$)					
	$\lambda = 0.5$	$\lambda = 0.8$	$\lambda = 1.0$	$\lambda = 1.2$	$\lambda = 2.0$
$H(Y)$ in bit	1.339	1.700	???	2.236	2.459
$D(P_X \|\| P_Y)$ in bit	0.2972	0.0515	???	0.0437	0.4610
$H(X)$ in bit	1.793				
Gegebene Binomialverteilung X					

© 2014 www.LNTwww.de

Die Spalten für $\lambda = 1$ sind in den Teilaufgaben (c) und (d) zu ergänzen. In der Teilaufgabe (f) sollen diese Ergebnisse interpretriert werden.

Hinweis: Die Aufgabe gehört zu **Kapitel 3.1.** Um die numerischen Berechnungen in Grenzen zu halten, werden folgende Hilfsgrößen vorgegeben; „lg” bezeichnet den Logarithmus zur Basis 10:

$$\begin{aligned} A' &= 0.4096 \cdot \lg \frac{0.4096}{0.3679} + 0.2048 \cdot \lg \frac{0.2048}{0.1839} + 0.0512 \cdot \lg \frac{0.0512}{0.0613} + \\ &+ 0.0064 \cdot \lg \frac{0.0064}{0.0153} + 0.0003 \cdot \lg \frac{0.0003}{0.0031} = \underline{= 0.021944}\,, \\ B' &= 0.1839 \cdot \lg\,(0.1839) + 0.0613 \cdot \lg\,(0.0613) + 0.0153 \cdot \lg\,(0.0153) + \\ &+ 0.0031 \cdot \lg\,(0.0031) + 0.0005 \cdot \lg\,(0.0005) + 0.0001 \cdot \lg\,(0.0001) \;\underline{= -0.24717}\,. \end{aligned}$$

Fragebogen zu "A3.4: KLD & Binomialverteilung"

a) Wie lauten die Kenngrößen der vorliegenden Binomialverteilung?
Hinweis: Geben Sie (maximal) eine Nachkommastelle ein.

$\boldsymbol{I}$ =

$\boldsymbol{p}$ =

$\boldsymbol{m_x}$ =

$\boldsymbol{\sigma_x}^2$ =

b) Welche Kullback–Leibler–Distanz sollte man hier verwenden?

- Keine der beiden Distanzen ist anwendbar.
- $D(P_X||P_Y)$ ist besser geeignet.
- $D(P_Y||P_X)$ ist besser geeignet.
- Beide Kullback–Leibler–Distanzen sind anwendbar.

c) Berechnen Sie die geeignete Kullback–Leibler–Distanz (hier mit D bezeichnet) für $\lambda = 1$. Berücksichtigen Sie die angegebene Hilfsgröße A'.

$\boldsymbol{\lambda = 1}$: $\boldsymbol{D}$ = bit

d) Berechnen Sie die Entropie $H(Y)$ der Poisson–Näherung mit der Rate $\lambda = 1$. Berücksichtigen Sie die angegebene Hilfsgröße B'.

$\boldsymbol{\lambda = 1}$: $\boldsymbol{H(Y)}$ = bit

e) Welche der folgenden Aussagen sind zutreffend?

- ☐ $H(Y)$–Berechnung: Alle Terme haben gleiches Vorzeichen.
- ☐ $D(P_X||P_Y)$–Berechnung: Alle Terme haben gleiches Vorzeichen.

f) Wie interpretieren Sie die vervollständigte Ergebnistabelle?

- ☐ Nach der Kullback–Leibler–Distanz sollte man $\lambda = 1$ wählen.
- ☐ $\lambda = 1$ garantiert auch die beste Approximation $H(Y) \approx H(X)$.

Z3.4: Nochmals KL–Distanz

N	$P_Y(1)$	$P_Y(2)$	$P_Y(3)$	$P_Y(4)$	$H(Y)$ (in bit)	$D(P_X \| P_Y)$ (in bit)
4	0.0000	0.2500	0.5000	0.2500	1.5000	???
10	0.5000	0.1000	0.3000	0.1000	???	???
40	0.2750	0.1750	0.2250	0.3250	1.9634	$3.76 \cdot 10^{-2}$
100	0.2400	0.1600	0.3000	0.3000	???	???
400	0.2550	0.2450	0.2675	0.2325	1.9981	$1.92 \cdot 10^{-3}$
1 000	0.2250	0.2530	0.2500	0.2720	???	???
4 000	0.2560	0.2463	0.2398	0.2580	1.9994	$4.87 \cdot 10^{-4}$
10 000	0.2471	0.2570	0.2548	0.2411	1.9995	$4.63 \cdot 10^{-4}$
40 000	0.2448	0.2529	0.2489	0.2534	1.9999	$1.40 \cdot 10^{-4}$
100 000	0.2518	0.2511	0.2492	0.2479	2.0000	$2.74 \cdot 10^{-5}$
400 000	0.2488	0.2504	0.2492	0.2516	2.0000	$1.38 \cdot 10^{-5}$

© 2014 www.LNTwww.de

Die Wahrscheinlichkeitsfunktion lautet:

$$P_Y(X) = [0.25, 0.25, 0.25, 0.25]\,.$$

Die Zufallsgröße X ist also gekennzeichnet

- durch den Symbolumfang $M = 4$,
- mit gleichen Wahrscheinlichkeiten.

Die Zufallsgröße Y ist stets eine Näherung für X. Sie wurde per Simulation aus einer Gleichverteilung gewonnen, wobei jeweils nur N Zufallswerte ausgewertet wurden. Das heißt: $P_Y(1)$, ... , $P_Y(4)$ sind im herkömmlichen Sinn keine Wahrscheinlichkeiten. Sie beschreiben vielmehr **relative Häufigkeiten.**

Das Ergebnis der sechsten Versuchsreihe (mit N = 1000) wird demnach durch die folgende Wahrscheinlichkeitsfunktion zusammengefasst:

$$P_Y(X) = [0.225\,, 0.253\,, 0.250\,, 0.272]\,.$$

Bei dieser Schreibweise ist bereits berücksichtigt, dass die Zufallsgrößen X und Y auf dem gleichen Alphabet $X = \{1, 2, 3, 4\}$ basieren.

Mit diesen Voraussetzungen gilt für die relative Entropie (englisch: *Informational Divergence*) zwischen den Wahrscheinlichkeitsfunktionen $P_X(\cdot)$ und $P_Y(\cdot)$:

$$D(P_X \,||\, P_Y) = {\rm E}_X \left[\log_2 \frac{P_X(X)}{P_Y(X)}\right] = \sum_{\mu=1}^{M} P_X(\mu) \cdot \log_2 \frac{P_X(\mu)}{P_Y(\mu)}\,.$$

Man bezeichnet $D(P_X||P_Y)$ als *Kullback–Leibler–Distanz*. Diese ist ein Maß für die Ähnlichkeit zwischen den beiden Wahrscheinlichkeitsfunktionen $P_X(\cdot)$ und $P_Y(\cdot)$. Die Erwartungswertbildung geschieht hier hinsichtlich der (tatsächlich gleichverteilten) Zufallsgröße X. Dies wird durch die Nomenklatur $E_X[\cdot]$ angedeutet.

Eine zweite Form der Kullback–Leibler–Distanz ergibt sich durch die Erwartungswertbildung hinsichtlich der Zufallsgröße Y $\Rightarrow$ $E_Y[\cdot]$:

$$D(P_Y \,||\, P_X) = {\rm E}_Y \left[\log_2 \frac{P_Y(X)}{P_X(X)}\right] = \sum_{\mu=1}^{M} P_Y(\mu) \cdot \log_2 \frac{P_Y(\mu)}{P_X(\mu)}\,.$$

Hinweis: Die Aufgabe bezieht sich auf das **Kapitel 3.1** dieses Buches. Die Angaben der Entropie $H(Y)$ und der Kullback–Leibler–Distanz $D(P_X||P_Y)$ in obiger Grafik sind in „bit” zu verstehen. Die mit „**???**” versehenen Felder sollen von Ihnen in dieser Aufgabe ergänzt werden.

Fragebogen zu "Z3.4: Nochmals KL–Distanz"

a) Welche Entropie besitzt die Zufallsgröße X?

$H(X)$ = bit

b) Wie groß sind die Entropien der Zufallsgrößen Y (Näherungen für X)?

$N = 1000$: $H(Y)$ = bit

$N = 100$: $H(Y)$ = bit

$N = 10$: $H(Y)$ = bit

c) Berechnen Sie die folgenden Kullback–Leibler–Distanzen.

$N = 1000$: $D(P_X || P_Y)$ = bit

$N = 100$: $D(P_X || P_Y)$ = bit

$N = 10$: $D(P_X || P_Y)$ = bit

d) Liefert $D(P_Y || P_X)$ jeweils exakt das gleiche Ergebnis?

- ○ Ja.
- ○ Nein.

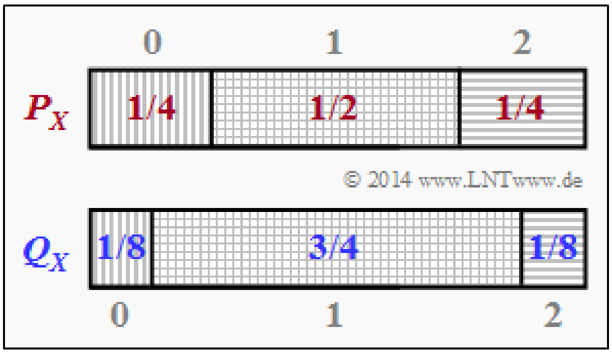
e) Welche Aussagen gelten für die Kullback–Leibler–Distanzen bei $N = 4$?

- ☐ Es gilt $D(P_X || P_Y) = 0$.
- ☐ Es gilt $D(P_X || P_Y) = 0.5$ bit.
- ☐ $D(P_X || P_Y)$ ist unendlich groß.
- ☐ Es gilt $D(P_Y || P_X) = 0$.
- ☐ Es gilt $D(P_Y || P_X) = 0.5$ bit.
- ☐ $D(P_Y || P_X)$ ist unendlich groß.

f) Ändern sich $H(Y)$ und $D(P_X || P_Y)$ monoton mit N?

- ○ Ja.
- ○ Nein.

A3.5: Partitionierungsungleichung



Die **Kullback–Leibler–Distanz** (kurz KLD) wird auch in der „Partitionierungsungleichung” (englisch: *Partition Unequality*) verwendet:

- Wir gehen von der Menge
$$X = \{\, x_1,\, x_2, \ldots,\, x_M \,\}$$
und den Wahrscheinlichkeitsfunktionen
$$P_X(X) \;=\; P_X(x_1, x_2, \ldots, x_M)\,,$$
$$Q_X(X) \;=\; Q_X(x_1, x_2, \ldots, x_M)$$
aus, die in irgendeiner Form „ähnlich” sein sollen.
- Die Menge X unterteilen wir in die Partitionen A_1, ..., A_K, die zueinander **disjunkt** sind und ein **vollständiges System** ergeben:
$$\bigcup_{i=1}^{K} A_i = X\,, \qquad A_i \cap A_j = \phi \text{ für } 1 \le i \ne j \le K\,.$$
- Die Wahrscheinlichkeitsfunktionen bezüglich der Partitionierungen $A = \{A_1, A_2, \ldots, A_K\}$ bezeichnen wir im Folgenden mit
$$P_X^{(A)} \;=\; [P_X(A_1)\,, \ldots\,, P_X(A_K)]\,, \qquad \text{wobei } P_X(A_i) = \sum_{x \in A_i} P_X(x)\,,$$
$$Q_X^{(A)} \;=\; [Q_X(A_1)\,, \ldots\,, Q_X(A_K)]\,, \qquad \text{wobei } Q_X(A_i) = \sum_{x \in A_i} Q_X(x)\,.$$

Die **Partitionierungsungleichung** liefert folgende Größenrelation hinsichtlich der Kullback–Leibler–Distanzen:
$$D(P_X^{(A)} \,||\, Q_X^{(A)}) \;\;\le\;\; D(P_X \,||\, Q_X)\,.$$

In der Aufgabe (a) soll die Kullback–Leibler–Distanz der beiden Wahrscheinlichkietsfunktionen $P_X(X)$ und $Q_X(X)$ für $X = \{0, 1, 2\}$ ⇒ $|X| = 3$ ermittelt werden. Anschließend soll die Menge X entsprechend

- $A = \{A_1, A_2\}$ mit $A_1 = \{0\}$ und $A_2 = \{1, 2\}$,
- $B = \{B_1, B_2\}$ mit $B_1 = \{1\}$ und $B_2 = \{0, 2\}$,
- $C = \{C_1, C_2\}$ mit $C_1 = \{2\}$ und $C_2 = \{0, 1\}$

mit $K = 2$ partitioniert werden und es sollen die jeweiligen Kullback–Leibler–Distanzen

- $D(P_X^{(A)}) \,||\, Q_X^{(A)})$,
- $D(P_X^{(B)}) \,||\, Q_X^{(B)})$,
- $D(P_X^{(C)}) \,||\, Q_X^{(C)})$

angegeben werden. In Aufgabe (e) wird schließlich nach den Bedingungen gefragt, damit in der obigen

Ungleichung das Gleichheitszeichen zutrifft.

Hinweis: Die Aufgabe gehört zu **Kapitel 3.1.** Die Wahrscheinlichkeitsfunktionen können aus obiger Grafik abgelesen werden:

$$P_X(X) = [1/4\,,\, 1/2,\, 1/4]\,,$$
$$Q_X(X) = [1/8\,,\, 3/4,\, 1/8]\,.$$

Fragebogen zu "A3.5: Partitionierungsungleichung"

a) Berechnen Sie die Kullback–Leibler–Distanz (KLD) allgemein.

$\boldsymbol{D(P_x \,\|\, Q_x)}$ = (bit)

b) Welche KLD ergibt sich für die Partitionierung $A_1 = \{0\}, A_2 = \{1, 2\}$?

$\boldsymbol{D(P_x{}^{(A)} \,\|\, Q_x{}^{(A)})}$ = (bit)

c) Welche KLD ergibt sich für die Partitionierung $B_1 = \{1\}, B_2 = \{0, 2\}$?

$\boldsymbol{D(P_x{}^{(B)} \,\|\, Q_x{}^{(B)})}$ = (bit)

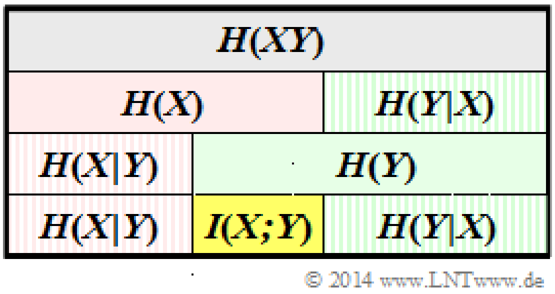
d) Welche KLD ergibt sich für die Partitionierung $C_1 = \{2\}, C_2 = \{0, 1\}$?

- ☐ Das gleiche Ergebnis wie für die Partitionierung A.
- ☐ Das gleiche Ergebnis wie für die Partitionierung B.
- ☐ Ein ganz anderes Ergebnis.

e) Unter welchen Bedingungen ergibt sich für allgemeines K die Gleichheit?

- ☐ Es müssen $|X|$ Gleichungen erfüllt sein.
- ☐ Für $x \in A_i$ muss gelten: $P_X(x)/Q_X(x) = P_X(A_i)/Q_X(A_i)$.

A3.6: Einige Entropieberechnungen



Wir betrachten die beiden Zufallsgrößen *XY* und *UV* mit den folgenden 2D–Wahrscheinlichkeitsfunktionen:

$$P_{XY}(X,Y) = \begin{pmatrix} 0.18 & 0.16 \\ 0.02 & 0.64 \end{pmatrix},$$

$$P_{UV}(U,V) = \begin{pmatrix} 0.068 & 0.132 \\ 0.272 & 0.528 \end{pmatrix}.$$

Für die Zufallsgröße *XY* sollen in dieser Aufgabe berechnet werden:

- die Verbundentropie (englisch: *Joint Entropy*):

$$H(XY) = -\mathrm{E}\left[\log_2 P_{XY}(X,Y)\right],$$

- die beiden Einzelentropien:

$$H(X) = -\mathrm{E}\left[\log_2 P_X(X)\right],$$
$$H(Y) = -\mathrm{E}\left[\log_2 P_Y(Y)\right].$$

Daraus lassen sich entsprechend dem obigen Schema – dargestellt für die Zufallsgröße *XY* – noch die folgenden Beschreibungsgrößen sehr einfach bestimmen:

- die bedingten Entropien (englisch: *Conditional Entropies*):

$$H(X|Y) = -\mathrm{E}\left[\log_2 P_{X|Y}(X|Y)\right],$$
$$H(Y|X) = -\mathrm{E}\left[\log_2 P_{Y|X}(Y|X)\right].$$

- die Transinformation (englisch: *Mutual Information*) zwischen *X* und *Y*:

$$I(X;Y) = \mathrm{E}\left[\log_2 \frac{P_{XY}(X,Y)}{P_X(X)\cdot P_Y(Y)}\right].$$

Abschließend sind qualitative Aussagen hinsichtlich der zweiten Zufallsgröße *UV* zu verifizieren.

Hinweis: Die Aufgabe bezieht sich auf das Themengebiet von **Kapitel 3.2.**

Fragebogen zu "A3.6: Einige Entropieberechnungen"

a) Berechnen Sie die Verbundentropie.

$\boldsymbol{H(XY)}$ = bit

b) Welche Entropien weisen die 1D–Zufallsgrößen X und Y auf?

$\boldsymbol{H(X)}$ = bit

$\boldsymbol{H(Y)}$ = bit

c) Wie groß ist die Transinformation zwischen den Zufallsgrößen X und Y?

$\boldsymbol{I(X;\, Y)}$ =

d) Berechnen Sie die beiden bedingten Entropien.

$\boldsymbol{H(X|Y)}$ = bit

$\boldsymbol{H(Y|X)}$ = bit

e) Welche der folgenden Aussagen treffen für die 2D–Zufallsgröße UV zu?

- ☐ Die 1D–Zufallsgrößen U und V sind statistisch unabhängig.
- ☐ Die gemeinsame Information von U und V $\Rightarrow$ $I(U;\, V)$ ist 0.
- ☐ Für die Verbundentropie gilt $H(UV) = H(XY)$.
- ☐ Es gelten die Beziehungen $H(U|V) = H(U)$ und $H(V|U) = H(V)$.

A3.7: Nochmals Transinformation

Wir betrachten das Tupel $Z = (X, Y)$, wobei die Einzelkomponenten X und Y jeweils ternäre Zufallsgrößen darstellen:

$$X = \{0, 1, 2\}\,, \quad Y = \{0, 1, 2\}\,.$$

$P_{XY}(X, Y)$

Y \ X	0	1	2
0	1/9	1/9	1/9
1	1/9	1/9	1/9
2	1/9	1/9	1/9

$P_{XW}(X, W)$

W \ X	0	1	2
0	1/9		
1	1/9	1/9	
2	1/9	1/9	1/9
3		1/9	1/9
4			1/9

© 2014 www.LNTwww.de

Die gemeinsame Wahrscheinlichkeitsfunktion $P_{XY}(X, Y)$ beider Zufallsgrößen ist oben angegeben. In der **Zusatzaufgabe Z3.7** wird diese Konstellation ausführlich analysiert. Man erhält als Ergebnis:

- $H(X) = H(Y) = \log_2(3) = 1.585$ bit,
- $H(XY) = \log_2(9) = 3.170$ bit,
- $I(X, Y) = 0$,
- $H(Z) = H(XZ) = 3.170$ bit,
- $I(X, Z) = 1.585$ bit.

Desweiteren betrachten wir hier die Zufallsgröße $W = \{0, 1, 2, 3, 4\}$, deren Eigenschaften sich aus der Verbundwahrscheinlichkeitsfunktion $P_{XW}(X, W)$ nach der unteren Skizze ergeben. Die Wahrscheinlichkeiten in allen weiß hinterlegten Feldern sind jeweils 0.

Gesucht ist in der vorliegenden Aufgabe die Transinformation

- zwischen den Zufallsgrößen X und W $\Rightarrow$ $I(X; W)$,
- zwischen den Zufallsgrößen Z und W $\Rightarrow$ $I(Z; W)$.

Hinweis: Die Aufgabe bezieht sich auf **Kapitel 3.2.**

Fragebogen zu "A3.7: Nochmals Transinformation"

a) Wie könnten die Größen X, Y und W zusammenhängen? Es gilt

- [ ] $W = X + Y$,
- [ ] $W = X - Y + 2$,
- [ ] $W = Y - X + 2$.

b) Welche Transinformation besteht zwischen den Zufallsgrößen X und W?

$\boldsymbol{I(X;\, W)}$ = bit

c) Welche Transinformation besteht zwischen den Zufallsgrößen Z und W?

$\boldsymbol{I(Z;\, W)}$ = bit

d) Welche der nachfolgenden Aussagen sind zutreffend?

- [ ] Es gilt $H(ZW) = H(XW)$.
- [ ] Es gilt $H(W|Z) = 0$.
- [ ] Es gilt $I(Z;\, W) > I(X;\, W)$.

Z3.7: Tupel aus ternären Zufallsgrößen

Wir betrachten das Tupel $Z = (X, Y)$, wobei die Einzelkomponenten X und Y jeweils ternäre Zufallsgrößen darstellen ⇒ Symbolumfang $|X| = |Y| = 3$. Die gemeinsame Wahrscheinlichkeitsfunktion $P_{XY}(X, Y)$ ist rechts angegeben.

$P_{XY}(X, Y)$

Y \ X	0	1	2
0	1/9	1/9	1/9
1	1/9	1/9	1/9
2	1/9	1/9	1/9

© 2014 www.LNTwww.de

In dieser Aufgabe sind zu berechnen:

- die Verbundentropie $H(XY)$ und die Transinformation $I(X; Y)$,
- die Verbundentropie $H(XZ)$ und die Transinformation $I(X; Z)$,
- die bedingten Entropien $H(Z|X)$ und $H(X|Z)$.

Hinweis: Die Aufgabe bezieht sich auf das Themengebiet von **Kapitel 3.2.**

Fragebogen zu "Z3.7: Tupel aus ternären Zufallsgrößen"

a) Berechnen Sie die folgenden Entropien.

$H(X)$ = bit

$H(Y)$ = bit

$H(XY)$ = bit

b) Welche Transinformationen besteht zwischen den Zufallsgrößen X und Y?

$I(X;\, Y)$ = bit

c) Welche Transinformation besteht zwischen den Zufallsgrößen X und Z?

$I(X;\, Z)$ = bit

d) Welche bedingten Entropien bestehen zwischen X und Z?

$H(Z|X)$ = bit

$H(X|Z)$ = bit

A3.8: Bedingte Transinformation

X	Y	Z	W
1	1	1	2
1	2	1	3
2	1	1	3
2	2	1	4
1	1	2	4
1	2	2	6
2	1	2	6
2	2	2	8

© 2014 www.LNTwww.de

Wir gehen von den statistisch unabhängigen Zufallsgrößen X, Y und Z mit den folgenden Eigenschaften aus:

$$X \in \{1, 2\}\,, \qquad Y \in \{1, 2\}\,, \qquad Z \in \{1, 2\}\,,$$
$$P_X(X) = P_Y(Y) = \big[1/2,\, 1/2\big]\,, \quad P_Z(Z) = \big[p, 1-p\big]\,.$$

Aus X, Y und Z bilden wir die neue Zufallsgröße

$$W = (X+Y)\cdot Z\,.$$

Damit ist offensichtlich, dass es zwischen den beiden Zufallsgrößen X und W statistische Abhängigkeiten gibt, die sich auch in der Transinformation $I(X;\, W) \ne 0$ zeigen werden.

Außerdem wird auch $I(Y;\, W) \ne 0$ sowie $I(Z;\, W) \ne 0$ gelten, worauf in dieser Aufgabe jedoch nicht näher eingegangen wird.

In dieser Aufgabe werden drei verschiedene Transinformationsdefinitionen verwendet:

- die *herkömmliche* Transinformation zwischen X und W:

  $$I(X;W) = H(X) - H(X|W)\,,$$

- die *bedingte* Transinformation zwischen X und W bei *gegebenem Festwert* $Z = z$:

  $$I(X;W\,|Z=z) = H(X\,|Z=z) - H(X|W, Z=z)\,,$$

- die *bedingte* Transinformation zwischen X und W bei *gegebener Zufallsgröße* Z:

  $$I(X;W\,|Z) = H(X\,|Z) - H(X|W Z)\,.$$

Der Zusammenhang zwischen den beiden letzten Definitionen lautet:

$$I(X;W\,|Z) = \sum_{z\,\in\,\mathrm{supp}(P_Z)} P_Z(z)\cdot I(X;W\,|Z=z)\,.$$

Hinweis: Die Aufgabe gehört zum Themengebiet von **Kapitel 3.2.**

Fragebogen zu "A3.8: Bedingte Transinformation"

a) Wie groß ist die Transinformation zwischen X und W, falls stets $Z = 1$ gilt?

$I(X;\, W\,|\,Z = 1)$ = (bit)

b) Wie groß ist die Transinformation zwischen X und W, falls stets $Z = 2$ gilt?

$I(X;\, W\,|\,Z = 2)$ = (bit)

c) Nun gelte $p = \Pr(Z = 1)$. Wie groß ist die bedingte Transinformation zwischen X und W unter der Annahme, dass $z \in Z = \{1, 2\}$ bekannt ist?

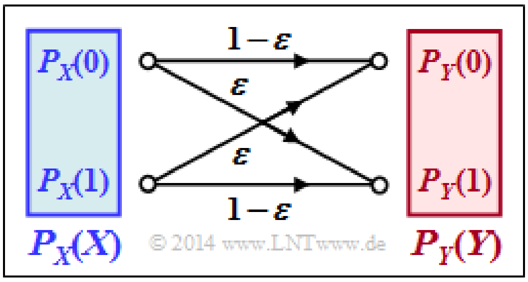
$p = 1/2$: $I(X;\, W\,|\,Z)$ = (bit)

$p = 3/4$: $I(X;\, W\,|\,Z)$ = (bit)

d) Wie groß ist die unkonditionierte Transinformation?

$p = 1/2$: $I(X;\, W)$ = (bit)

A3.9: Transinformation beim BSC



Wir betrachten den **Binary Symmetric Channel** (BSC). Für die gesamte Aufgabe gelten die Parameterwerte:

- Verfälschungswahrscheinlichkeit: $\varepsilon = 0.1$,
- Wahrscheinlichkeit für 0: $p_0 = 0.2$,
- Wahrscheinlichkeit für 1: $p_1 = 0.8$.

Damit lautet die Wahrscheinlichkeitsfunktion der Quelle:

$$P_X(X) = (0.2\,,\, 0.8)\,,$$

und für die Quellenentropie gilt:

$$H(X) = p_0 \cdot \log_2 \frac{1}{p_0} + p_1 \cdot \log_2 \frac{1}{p_1} = H_{\rm bin}(0.2) = 0.7219\,{\rm bit}\,.$$

In der Aufgabe sollen ermittelt werden:

- die Wahrscheinlichkeitsfunktion der Sinke:

$$P_Y(Y) = \big(\,P_Y(0)\,,\,P_Y(1)\,\big)\,,$$

- die Verbundwahrscheinlichkeitsfunktion:

$$P_{XY}(X,Y) = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}\,,$$

- die Transinformation

$$I(X;Y) = {\rm E}\left[\log_2 \frac{P_{XY}(X,Y)}{P_X(X) \cdot P_Y(Y)}\right]\,,$$

- die Äquivokation:

$$H(X|Y) = {\rm E}\left[\log_2 \frac{1}{P_{X|Y}(X\,|\,Y)}\right]\,,$$

- die Irrelevanz:

$$H(Y|X) = {\rm E}\left[\log_2 \frac{1}{P_{Y|X}(Y\,|\,X)}\right]\,.$$

Hinweis: Die Aufgabe gehört zu **Kapitel 3.3.** In der **Aufgabe Z3.9** wird die **Kanalkapazität** $C_{\rm BSC}$ des BSC–Modells berechnet. Diese ergibt sich als die maximale Transinformation $I(X;\ Y)$ durch Maximierung bezüglich der Symbolwahrscheinlichkeiten p_0 bzw. $p_1 = 1 - p_0$.

Fragebogen zu "A3.9: Transinformation beim BSC"

a) Berechnen Sie die Verbundwahrscheinlichkeiten $P_{XY}(X, Y)$.

$P_{XY}(0, 0)$ =

$P_{XY}(0, 1)$ =

$P_{XY}(1, 0)$ =

$P_{XY}(1, 1)$ =

b) Wie lautet die Wahrscheinlichkeitsfunktion $P_Y(Y)$?

$P_Y(0)$ =

$P_Y(1)$ =

c) Welcher Wert ergibt sich für die Transinformation?

$I(X; Y)$ = bit

d) Welcher Wert ergibt sich für die Äquivokation?

$H(X|Y)$ = bit

e) Welche Aussage trifft für die Sinkenentropie $H(Y)$ zu?

- $H(Y)$ ist nie größer als $H(X)$.
- $H(Y)$ ist nie kleiner als $H(X)$.

f) Welche Aussage trifft für die Irrelevanz $H(Y|X)$ zu?

- $H(Y|X)$ ist nie größer als die Äquivokation $H(X|Y)$.
- $H(Y|X)$ ist nie kleiner als die Äquivokation $H(X|Y)$.

Z3.9: BSC–Kanalkapazität

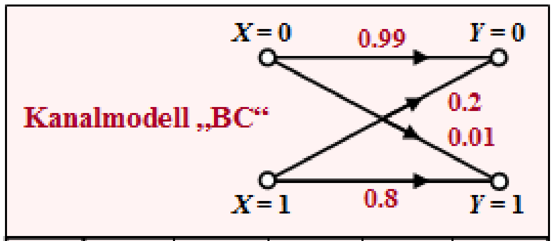
Die Kanalkapazität C wurde von Claude E. Shannon als die maximale Transinformation definiert, wobei sich die Maximierung allein auf die Quellenstatistik bezieht:

$$C = \max_{P_X(X)} \; I(X;Y)\,.$$

Beim Binärkanal mit der Wahrscheinlichkeitsfunktion $P_X(X) = [p_0,\; p_1]$ ist nur ein Parameter optimierbar, beispielsweise p_0. Die Wahrscheinlichkeit für eine „1” ist damit ebenfalls festgelegt: $p_1 = 1 - p_0$.

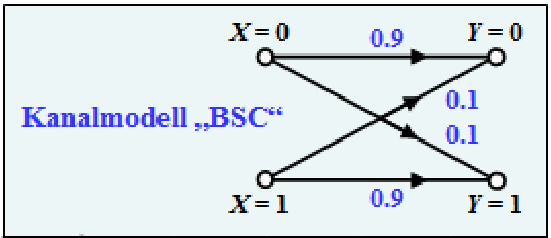
Die obere Grafik (rot hinterlegt) fasst die Ergebnisse für den **unsymmetrischen Binärkanal** mit $\varepsilon_0 = 0.01$ und $\varepsilon_1 = 0.2$ zusammen, der im Theorieteil betrachtet wurde. Die Maximierung führt zum Ergebnis $p_0 = 0.55$ ⇒ $p_1 = 0.45$, und man erhält für die Kanalkapazität:

$$C_{\rm BC} = \max_{P_X(X)} \; I(X;Y)\big|_{p_0 = 0.55} = 0.5779 \; {\rm bit}\,.$$



p_0	$H(X)$	$H(X \vert Y)$	$I(X; Y)$	$H(Y \vert X)$	$H(Y)$
0.40	0.9710	0.4372	0.5338	0.4655	0.9993
0.45	0.9928	0.4351	0.5577	0.4334	0.9911
0.50	1.0000	0.4276	0.5724	0.4014	0.9738
0.55	0.9928	0.4149	0.5779	0.3693	0.9472
0.60	0.9710	0.3974	0.5736	0.3372	0.9108

Alle Angaben in „bit“ © 2014 www.LNTwww.de



p_0	$H(X)$	$H(X \vert Y)$	$I(X; Y)$	$H(Y \vert X)$	$H(Y)$
0.40	0.9710	0.4585	0.5125	0.4690	0.9815
0.45	0.9928	0.4664	0.5264	0.4690	0.9954
0.50	1.0000	0.4690	0.5310	0.4690	1.0000
0.55	0.9928	0.4664	0.5264	0.4690	0.9954
0.60	0.9710	0.4585	0.5125	0.4690	0.9815

In der unteren Grafik (blaue Hinterlegung) sind die gleichen informationstheoretischen Größen für den symmetrischen Kanal ⇒ **Binary Symmetric Channel** (BSC) mit den Verfälschungswahrscheinlichkeiten $\varepsilon_1 = \varepsilon_2 = \varepsilon = 0.1$ angegeben, der auch für die **Aufgabe A3.9** vorausgesetzt wurde.

In der vorliegenden Aufgabe sollen Sie für das BSC–Kanalmodell (zunächst für $\varepsilon = 0.1$)

- die Entropien $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$ analysieren,
- den Quellenparameter p_0 hinsichtlich maximaler Transinformation $I(X; Y)$ optimieren,
- somit die Kanalkapazität $C(\varepsilon)$ bestimmen, sowie
- durch Verallgemeinerung eine geschlossene Gleichung für $C(\varepsilon)$ angeben.

Hinweis: Die Aufgabe bezieht sich auf die Thematik von **Kapitel 3.3.**

Fragebogen zu "Z3.9: BSC–Kanalkapazität"

a) Welche Aussagen gelten für die bedingten Entropien beim BSC?

- [ ] Die Äquivokation ergibt sich zu $H(X|Y) = H_{\rm bin}(\varepsilon)$.
- [ ] Die Irrelevanz ergibt sich zu $H(Y|X) = H_{\rm bin}(\varepsilon)$.
- [ ] Die Irrelevanz ergibt sich zu $H(Y|X) = H_{\rm bin}(p_0)$.

b) Welche Aussage gilt für die Kanalkapazität $C_{\rm BSC}$ des BSC–Modells?

- [ ] Die Kanalkapazität ist gleich der maximalen Transinformation.
- [ ] Die Maximierung führt beim BSC zum Ergebnis $p_0 = p_1 = 0.5$.
- [ ] Für $p_0 = p_1 = 1/2$ gilt $H(X) = H(Y) = 1$ bit.

c) Welche Kanalkapazität ergibt sich abhängig vom BSC–Parameter ε?

$\boldsymbol{\varepsilon = 0}$: $C_{\rm BSC}$ = (bit)

$\boldsymbol{\varepsilon = 0.1}$: $C_{\rm BSC}$ = (bit)

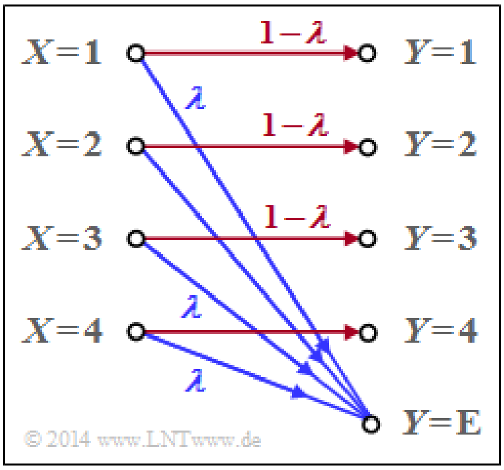
$\boldsymbol{\varepsilon = 0.5}$: $C_{\rm BSC}$ = (bit)

d) Welche Kanalkapazität ergibt sich abhängig vom BSC–Parameter ε?

$\boldsymbol{\varepsilon = 1}$: $C_{\rm BSC}$ = (bit)

$\boldsymbol{\varepsilon = 0.9}$: $C_{\rm BSC}$ = (bit)

A3.10: Auslöschungskanal



Betrachtet wird ein Auslöschungskanal mit

- den M Eingängen $x \in X = \{1, 2, \ldots, M\}$, und
- den $M + 1$ Ausgängen $y \in Y = \{1, 2, \ldots, M, \mathrm{E}\}$.

Die Grafik zeigt das Modell für den Sonderfall $M = 4$. Das Sinkensymbol $y = \mathrm{E}$ berücksichtigt eine Auslöschung (englisch: *Erasure*) für den Fall, dass der Empfänger keine hinreichend gesicherte Entscheidung treffen kann.

Die Übergangswahrscheinlichkeiten sind für $1 \le \mu \le M$ wie folgt gegeben:

$$\Pr(Y = \mu \,|\, X = \mu) \;=\; 1 - \lambda\,,$$
$$\Pr(Y = \mathrm{E} \,|\, X = \mu) \;=\; \lambda\,.$$

Gesucht ist

- die Kapazität $C_{M\text{–EC}}$ dieses *M–ary Erasure Channels*,
- die Kapazität $C_{\rm BEC}$ des **Binary Erasure Channels** als Sonderfall des obigen Modells.

Hinweis: Die Aufgabe beschreibt die Thematik von **Kapitel 3.3.** In dem obigen Schaubild sind Auslöschungen (Wahrscheinlichkeit λ) blau gezeichnet und „richtige Übertragungswege” (also von $X = \mu$ nach $Y = \mu$) blau ($1 \le \mu \le M$).

Fragebogen zu "A3.10: Auslöschungskanal"

a) Welches $P_X(X)$ ist zur Kanalkapazitätsberechnung allgemein anzusetzen?

- ☐ $P_X(X) = (0.5, 0.5)$,
- ☐ $P_X(X) = (1/M, 1/M, \ldots , 1/M)$,
- ☐ $P_X(X) = (0.1, 0.2, 0.3, 0.4)$.

b) Wie viele Wahrscheinlichkeiten $p_{\mu\kappa} = \Pr[(X = \mu) \cap (Y = \kappa)]$ sind $\neq 0$?

- ☐ Genau $M \cdot (M + 1)$,
- ☐ Genau M,
- ☐ Genau $2 \cdot M$.

c) Wie groß ist die Sinkenentropie allgemein und für $M = 4$ und $\lambda = 0.2$?

$\boldsymbol{M = 4, \lambda = 0.2:}$ $\boldsymbol{H(Y)}$ = bit

d) Berechnen Sie die Irrelevanz. Welcher Wert ergibt sich für $M = 4$, $\lambda = 0.2$?

$\boldsymbol{M = 4, \lambda = 0.2:}$ $\boldsymbol{H(Y|X)}$ = bit

e) Wie groß ist die Kanalkapazität in Abhängigkeit von M?

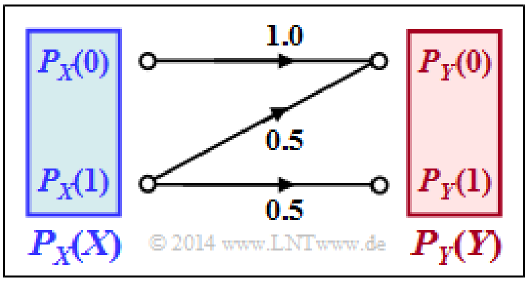
$\boldsymbol{M = 4:}$ $\boldsymbol{C}$ = bit

$\boldsymbol{M = 2:}$ $\boldsymbol{C}$ = bit

f) Wie lautet die Kanalkapazität des BEC–Kanals in kompakter Form?

- ☐ $C_{\rm BEC} = 1 - \lambda$,
- ☐ $C_{\rm BEC} = 1 - H_{\rm bin}(\lambda)$.

Z3.10: Extrem unsymmetrischer Kanal



Betrachtet wird der nebenstehend gezeichnete Kanal mit den folgenden Eigenschaften:

- Das Symbol $X = 0$ wird immer richtig übertragen und führt stets zum Ergebnis $Y = 0$.
- Das Symbol $X = 1$ wird maximal verfälscht. Aus Sicht der Informationstheorie bedeutet diese Aussage:

$$\Pr(Y = 0\,|\,X = 1) = \Pr(Y = 1\,|\,X = 1) = 0.5\,.$$

Zu bestimmen sind in dieser Aufgabe:

- die Transinformation $I(X;\,Y)$ für $P_X(0) = p_0 = 0.4$ und $P_X(1) = p_1 = 0.6$. Es gilt allgemein:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y)\,,\\ I(X;Y) &= H(Y) - H(Y|X)\,,\\ I(X;Y) &= H(X) + H(Y) - H(XY)\,. \end{aligned}$$

- die Kanalkapazität:

$$C = \max_{P_X(X)} \; I(X;Y)\,.$$

Hinweis: Die Aufgabe beschreibt einen Teilaspekt von **Kapitel 3.3.** In der **Aufgabe A3.13** sollen die hier gefundenen Ergebnisse im Vergleich zum BSC–Kanal interpretiert werden.

Fragebogen zu "Z3.10: Extrem unsymmetrischer Kanal"

a) Berechnen Sie die Quellenentropie allgemein und für $p_0 = 0.4$.

p_0 = 0.4: $H(X)$ = bit

b) Berechnen Sie die Sinkenentropie allgemein und für $p_0 = 0.4$.

p_0 = 0.4: $H(Y)$ = bit

c) Berechnen Sie die Verbundentropie allgemein und für $p_0 = 0.4$.

p_0 = 0.4: $H(XY)$ = bit

d) Berechnen Sie die Transinformation allgemein und für $p_0 = 0.4$.

p_0 = 0.4: $I(X; Y)$ = bit

e) Welche Wahrscheinlichkeit p_0 führt zur Kanalkapazität C?

Maximierung: p_0 =

f) Wie groß ist die Kanalkapazität des vorliegenden Kanals?

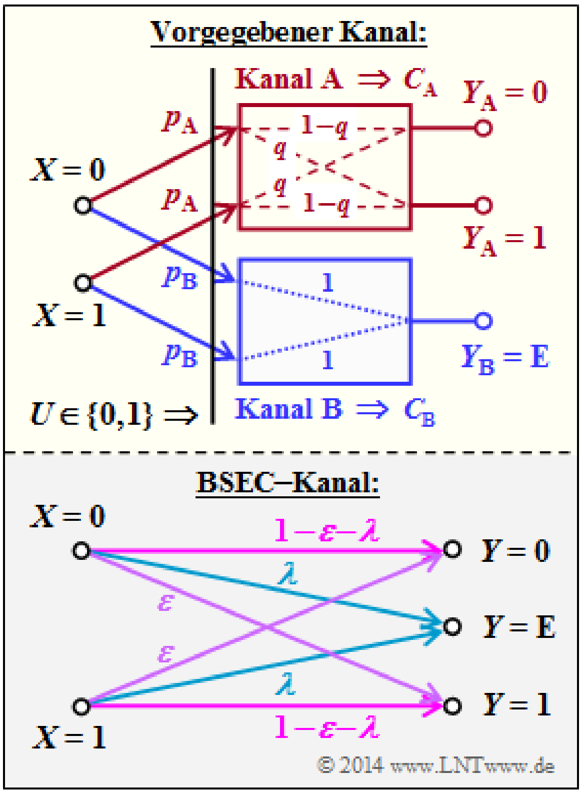
C = bit

g) Wie groß sind die bedingten Entropien?

p_0 gemäß (e): $H(X|Y)$ = bit

$H(Y|X)$ = bit

A3.11: Streng symmetrische Kanäle



Die obere Grafik zeigt zwei streng symmetrische Teilkanäle A und B. Ein **streng symmetrischer Kanal** (englisch: *Strongly Symmetric Channel*) ist dabei

- gleichmäßig **dispersiv** (*uniformly dispersive*) ⇒ jedes Eingangssymbol u hat die gleiche Menge an Übergangswahrscheinlichkeiten:
$$\{P_{Y|U}(y|u): \;\; u \in U\},$$
- zudem gleichmäßig **fokussierend** (*uniformly focusing*) ⇒ jedes Ausgangssymbol y hat die gleiche Übergangswahrscheinlichkeitsmenge:
$$\{P_{Y|U}(y|u): \;\; y \in Y\}.$$

Die Zufallsgröße $U = \{0, 1\}$ tritt dabei direkt an den Eingängen der Teilkanäle A und B auf.

Die Kanalkapazität eines streng symmetrischen Kanals lässt sich sehr viel einfacher berechnen als im unsymmetrischen Fall. Hierauf wird jedoch in dieser Aufgabe nicht näher eingegangen.

Für die Kapazität des Gesamtkanals gilt
$$C = p_A \cdot C_A + p_B \cdot C_B.$$

Hierbei bezeichnet p_A die Wahrscheinlichkeit, dass der Teilkanal A ausgewählt wird und C_A dessen Kapazität. Entsprechendes gilt für den Teilkanal B.

Anschließend soll auch die Kanalkapazität des **Binary Symmetric Error & Erasure Channel** (BSEC) nach der unteren Skizze (grau hinterlegt) ermittelt werden, indem der Zusammenhang hergeleitet wird zwischen

- den Parametern p_A, p_B und q des oben dargestelltern Teilkanalmodells, und
- den Parametern λ und ε des BSEC–Modells.

Hinweis: Die Aufgabe gehört zum Themengebiet von **Kapitel 3.3.** Entsprechend der **Aufgabe Z3.9** gilt für die Kanalkapazität des BSC–Modells mit der Verfälschungswahrscheinlichkeit ε:
$$C_{BSC} = 1 - H_{bin}(\varepsilon).$$

Fragebogen zu "A3.11: Streng symmetrische Kanäle"

a) Welche Kapazität C_A besitzt der Teilkanal A?

- ☐ $C_\text{A} = 1 - H_\text{bin}(q)$.
- ☐ $C_\text{A} = p_\text{A} \cdot [1 - H_\text{bin}(q)]$.
- ☐ $C_\text{A} = 0$.

b) Welche Kapazität C_B besitzt der Teilkanal B?

- ☐ $C_\text{B} = 1 - H_\text{bin}(q)$.
- ☐ $C_\text{B} = p_\text{B} \cdot [1 - H_\text{bin}(q)]$.
- ☐ $C_\text{B} = 0$.

c) Welche Kapazität C besitzt der Gesamtkanal?

- ☐ $C = 1 - H_\text{bin}(q)$.
- ☐ $C = p_\text{A} \cdot [1 - H_\text{bin}(q)]$.
- ☐ $C = 0$.

d) Wie gelangt man vom betrachteten Teilkanalmodell zum BSEC? Mit

- ☐ $p_\text{A} = \lambda$,
- ☐ $p_\text{A} = 1 - \lambda$,
- ☐ $p_\text{A} = \varepsilon$,
- ☐ $p_\text{A} = \varepsilon/(1 - \lambda)$?

e) Wie gelangt man vom betrachteten Teilkanalmodell zum BSEC? Mit

- ☐ $q = \lambda$,
- ☐ $q = 1 - \lambda$,
- ☐ $q = \varepsilon$,
- ☐ $q = \varepsilon/(1 - \lambda)$?

f) Berechnen Sie die BSEC–Kanalkapazität für $\varepsilon = 0.08$ und $\lambda = 0.2$.

$\boldsymbol{\varepsilon = 0.08, \lambda = 0.2}$: $\boldsymbol{C_\text{BSEC}}$ = ______ bit

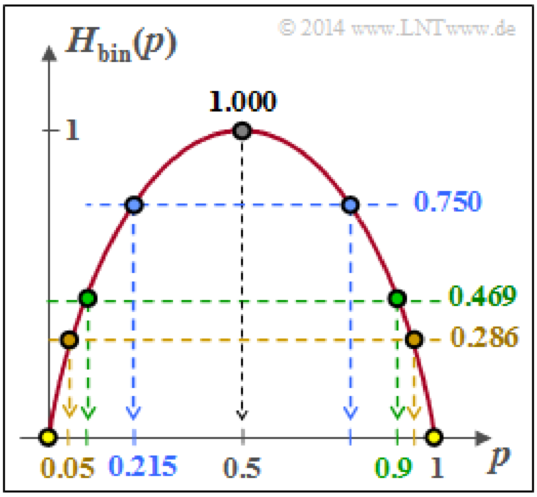
g) Wie groß ist die Kanalkapazität des BSC–Kanals für $\varepsilon = 0.08$?

$\boldsymbol{\varepsilon = 0.08}$: $C_{\rm BSC}$ = bit

h) Wie groß ist die Kanalkapazität des BEC–Kanals für $\lambda = 0.2$?

$\boldsymbol{\lambda = 0.2}$: $C_{\rm BEC}$ = bit

A3.12: Coderate und Zuverlässigkeit



Das Kanalcodierungstheorem von Shannon besagt unter anderem, dass über einen Kanal mit beliebig kleiner Blockfehlerwahrscheinlichkeit übertragen werden kann, so lange die Coderate R nicht größer ist als die Kanalkapazität C. Dieses Ergebnis erreicht man mit *Kanalcodierung* (englisch: *Channel Coding*) bei sehr großen Blocklängen: $n \to \infty$, was mit einem beliebig großen Aufwand verbunden ist.

Diese Aussage basiert auf informationstheoretischen Gesetzmäßigkeiten, die Shannon selbst aufgestellt hat. Shannon sagt nicht, wie diese „unendlich gute" Kanalcodierung aussehen muss, er beweist nur, dass es sie gibt.

Der Umkehrschluss des Kanalcodierungstheorems sagt aus, dass mit einer Rate $R > C$ keine fehlerfreie Übertragung möglich ist. Es gibt kein Codiersystem mit verschwindend kleiner Fehlerwahrscheinlichkeit, auch wenn die Codierung noch so aufwändig und ausgeklügelt wäre.

Vielmehr lassen sich für den Fall $R > C$ zwei Schranken angeben:

- Überträgt man über einen *diskreten gedächtnislosen Kanal* (DMC) mit einer Rate R größer als die Kanalkapazität C, so gibt es eine Schranke für die Bitfehlerwahrscheinlichkeit:

$$p_{\rm B} \ge H_{\rm bin}^{-1}\left(1 - \frac{C}{R}\right) > 0\,.$$

- Soll die Bitfehlerwahrscheinlichkeit nach bestmöglicher Decodierung den Wert $p_{\rm B}$ nicht überschreiten, so darf die Coderate einen vorgegebenen Grenzwert nicht unterschreiten:

$$R \le \frac{C}{1 - H_{\rm bin}(p_{\rm B})}\,.$$

In dieser Aufgabe sollen diese Gleichungen unter Verwendung der **binären Entropiefunktion**

$$H_{\rm bin}(p) = p \cdot \log_2 \frac{1}{p} + (1-p) \cdot \log_2 \frac{1}{1-p}\,.$$

numerisch ausgewertet werden. Diese ist oben skizziert. Wegen $0 < p_{\rm B} \le 1$ und $0 < C/R < 1$ liegt das Argument der Funktion $H_{\rm bin}(\cdot)$ und deren Umkehrfunktion $H_{\rm bin}{}^{-1}(\cdot)$ stets zwischen 0 und 1.

Hinweis: Die vorliegende Aufgabe behandelt einen Teilaspekt von **Kapitel 3.3.** Die Werte der binären Entropiefunktion werden zum Beispiel durch das folgende Flashmodul bereitgestellt:

Entropien von Nachrichtenquellen

Fragebogen zu "A3.12: Coderate und Zuverlässigkeit"

a) Wie lautet die untere Bitfehlerwahrscheinlichkeitsschranke für $R/C = 4$?

$\boldsymbol{R/C = 4}$: $\quad \boldsymbol{p_{\rm B,min}}$ =

b) Mit welcher Rate kann man über einen Kanal mit der Kanalkapazität $C = 0.5$ (bit) bei gegebener Grenz–Fehlerwahrscheinlichkeit übertragen?

$\boldsymbol{p_{\rm B} = 0.1}$: $\quad \boldsymbol{R_{\rm max}}$ =

$\boldsymbol{p_{\rm B} = 0.05}$: $\quad \boldsymbol{R_{\rm max}}$ =

c) Welche Relation besteht zwischen $H_{\rm bin}(p)$ und der Näherung $p \cdot \log_2 1/p$?

- [ ] Es gilt $H_{\rm bin}(p) < p \cdot \log_2 1/p$.
- [ ] Es gilt $H_{\rm bin}(p) > p \cdot \log_2 1/p$.
- [ ] $H_{\rm bin}(p) - p \cdot \log_2 1/p$ wird um so kleiner, je kleiner p ist.

d) Welche der folgenden Gleichungen sind obere Schranken für die Rate?

- [ ] $R/C \;\le\; [1 - H_{\rm bin}(p_{\rm B})]^{-1}$,
- [ ] $R'/C \;\le\; [1 + p_{\rm B} \cdot \log_2 (p_{\rm B})]^{-1}$,
- [ ] $R''/C \;\le\; 1 - p_{\rm B} \cdot \log_2 (p_{\rm B})$,
- [ ] $R'''/C \;\le\; 1 + i \cdot p_{\rm B}/ \lg(2)$ für $p_{\rm B} = 10^{-i}$.

A3.13: Kanalcodierungstheorem

Shannons Kanalcodierungstheorem besagt, dass über einen diskreten gedächtnislosen Kanal (DMC) mit der Coderate R fehlerfrei übertragen werden kann, so lange R nicht größer ist als die Kanalkapazität

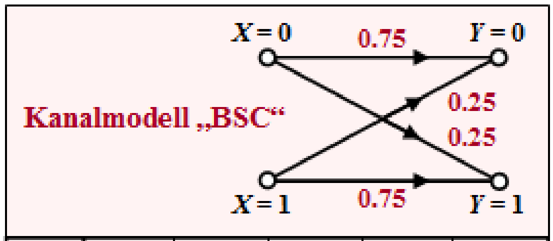
$$C = \max_{P_X(X)} \ I(X;Y).$$

Das Kanalcodierungstheorem soll in dieser Aufgabe numerisch ausgewertet werden, wobei zwei typische Kanalmodelle zu betrachten sind:

- das **BSC–Modell** (*Binary Symmetric Channel*) mit Verfälschungswahrscheinlichkeit ε = 0.25 und der Kanalkapazität $C = 1 - H_{\rm bin}(\varepsilon)$,
- das sog. **EUC–Modell** (*Extremely Unsymmetric Channel*) entsprechend der **Aufgabe Z3.10** (diese Bezeichnung ist nicht allgemein üblich).

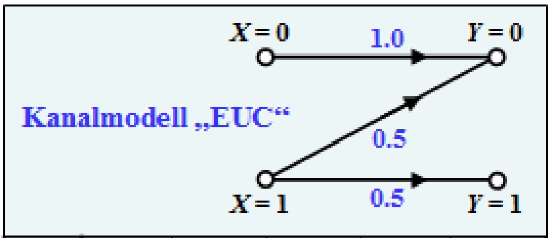
Hinweis: Diese Aufgabe gehört zum Themengebiet von **Kapitel 3.3.** Die Grafiken zeigen die numerischen Werte der informationstheoretischen Größen für die beiden Kanäle „BSC” und „EUC”:

- der Quellenentropie $H(X)$,
- der Äquivokation $H(X|Y)$,
- der Transinformation $I(X;Y)$,
- der Irrelevanz $H(Y|X)$, und
- der Sinkenentropie $H(Y)$.



p_0	$H(X)$	$H(X\|Y)$	$I(X;Y)$	$H(Y\|X)$	$H(Y)$
0.30	0.8813	0.7216	0.1597	0.8113	0.9710
0.40	0.9710	0.7895	0.1815	0.8113	0.9928
0.50	1.0000	0.8113	0.1887	0.8113	1.0000
0.60	0.9710	0.7895	0.1815	0.8113	0.9928
0.70	0.8813	0.7216	0.1597	0.8113	0.9710

Alle Angaben in „bit“ © 2014 www.LNTwww.de



p_0	$H(X)$	$H(X\|Y)$	$I(X;Y)$	$H(Y\|X)$	$H(Y)$
0.30	0.8813	0.6472	0.2341	0.7000	0.9341
0.40	0.9710	0.6897	0.2813	0.6000	0.8831
0.50	1.0000	0.6887	0.3113	0.5000	0.8113
0.60	0.9928	0.6490	0.3219	0.4000	0.7219
0.70	0.9710	0.5715	0.3098	0.3000	0.6098

Der Parameter in diesen Tabellen ist $p_0 = \Pr(X = 0)$ im Bereich von 0.3 bis 0.7. Entsprechend gilt für die Wahrscheinlichkeit des Quellensymbols „1”: $p_1 = 1 - p_0$.

Die (mittlere) Bitfehlerwahrscheinlichkeit wird in den Fragen mit $p_{\rm B}$ bezeichnet.

Fragebogen zu "A3.13: Kanalcodierungstheorem"

a) Welche Aussagen gelten für uncodierte Übertragung (R = 1), wenn man von $p_0 = p_1 = 0.5$ ausgeht?

- ☐ Mit BSC ergibt sich eine kleinere Bitfehlerwahrscheinlichkeit.
- ☐ Mit EUC ergibt sich eine kleinere Bitfehlerwahrscheinlichkeit.
- ☐ Beide Modelle führen zur gleichen Bitfehlerwahrscheinlichkeit.

b) Lässt sich bei $R = 1$ durch andere p_0, p_1 das Ergebnis (formal) verbessern?

- ☐ Bei beiden Kanälen.
- ☐ Beim BSC–Modell.
- ☐ Beim EUC–Modell.
- ☐ Bei keinem Modell.

c) Über welchen Kanal lässt sich mit der Rate $R = 0.16$ fehlerfrei übertragen?

- ☐ Bei beiden Kanälen.
- ☐ Beim BSC–Modell.
- ☐ Beim EUC–Modell.
- ☐ Bei keinem Modell.

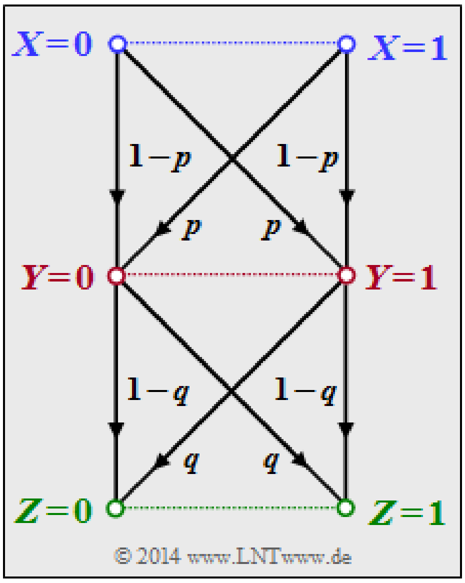
d) Über welchen Kanal lässt sich mit der Rate $R = 0.32$ fehlerfrei übertragen?

- ☐ Bei beiden Kanälen.
- ☐ Beim BSC–Modell.
- ☐ Beim EUC–Modell.
- ☐ Bei keinem Modell.

e) Über welchen Kanal lässt sich mit der Rate $R = 0.48$ fehlerfrei übertragen?

- ☐ Bei beiden Kanälen.
- ☐ Beim BSC–Modell.
- ☐ Beim EUC–Modell.
- ☐ Bei keinem Modell.

A3.14: Data Processing Theorem



Wir betrachten die folgende Datenverarbeitungskette:

- Binäre Eingangsdaten X werden durch den Prozessor 1 verarbeitet, der durch bedingte Wahrscheinlichkeiten ($P_{Y|X}$) beschreibbar ist. Dessen Ausgangsgröße ist Y.
- Ein zweiter Prozessor mit der Zufallsgröße Y am Eingang und der Zufallsgröße Z am Ausgang ist durch $P_{Z|Y}$ gegeben. Z hängt allein von Y ab (entweder deterministisch oder stochastisch) und ist unabhängig von X:

$$P_{Z|XY}(z|x, y) = P_{Z|Y}(z|y)\,.$$

Hierbei wurde folgende Nomenklatur benutzt:

$$x \in X = \{0, 1\}\,, \quad y \in Y = \{0, 1\}\,, \quad z \in Z = \{0, 1\}\,.$$

Die Verbund–Wahrscheinlichkeitsfunktion (englisch: *Joint Probability Mass Function*) lautet:

$$P_{XYZ}(x, y, z) = P_X(x) \cdot P_{Y|X}(y|x) \cdot P_{Z|Y}(z|y)\,.$$

Das bedeutet auch: $X \rightarrow Y \rightarrow Z$ bilden eine **Markovkette.** Für eine solche gilt das *Data Processing Theorem* mit folgender Konsequenz:

$$I(X;Z) \le I(X;Y)\,,$$
$$I(X;Z) \le I(Y;Z)\,.$$

Das Theorem besagt somit:

- Man kann durch Manipulation (*Processing*) der Daten Y keine zusätzliche Information über den Eingang X gewinnen.
- Datenverarbeitung (durch den Prozessor 2) dient nur dem Zweck, die Information über X besser sichtbar zu machen.

Hinweis: Die Aufgabe gehört zu **Kapitel 3.3.**

Fragebogen zu "A3.14: Data Processing Theorem"

a) Wie lässt sich das Ergebnis $I(X;\, Y) = 1 - H_{\rm bin}(p)$ herleiten?

- ☐ Über die Eigenschaften eines streng symmetrischen Kanals.
- ☐ Weil $H_{\rm bin}(p)$ eine konkave Funktion ist.
- ☐ Das Ergebnis gilt für jede Wahrscheinlichkeitsfunktion $P_X(X)$.

b) Welche Transinformation ergibt sich für den Prozessor 1 mit $p = 0.1$?

$\boldsymbol{p = 0.1}$: $\boldsymbol{I(X;\, Y)}$ = (bit)

c) Welche Transinformation ergibt sich für den Prozessor 2 mit $q = 0.2$?

$\boldsymbol{q = 0.2}$: $\boldsymbol{I(Y;\, Z)}$ = (bit)

d) Welche Transinformation ergibt sich für das Gesamtsystem?

$\boldsymbol{p = 0.1,\ q = 0.2}$: $\boldsymbol{I(X;\, Z)}$ = (bit)

e) Erfüllt dieses Beispiel das *Data Processing Theorem*?

- ○ Ja.
- ○ Nein.