2.5 *Grundlegende Theoreme der lnformationstheorie*

2.50 *Ein Hilfssatz: Die Jensensche Ungleichung.*

Mehrere grundlegende Theoreme der Informationstheorie lassen sich bequem mittels der Jensenschen Ungleichung (Bild 2.50a) beweisen. Diese besagt, daß bei einer (in einem lntervall) konvexen Funktion $f(x)$ der gewichtete Mittelwert $\overline{f}(x)$ der Funktionswerte stets kleiner ist als der Funktionswert $f(\overline{x})$ des gewichteten Mittelwerts $\overline{x}$ der zugehörigen Argumentwerte x_i, sofern diese nicht alle übereinstimmen. Dabei heißt eine Funktion $f(x)$ in einem Intervall $]a, b[$ **konvex** (genauer: „von oben" konvex), wenn - anschaulich gesprochen - für zwei beliebige, verschiedene Punkte x_1, x_2 dieses Intervalls die Strecke zwischen den Kurvenpunkten $f(x_1)$ und $f(x_2)$ ganz unterhalb der Kurve liegt, welche $f(x)$ darstellt – selbstverständlich mit Ausnahme des Anfangs- und des Endpunkts dieser Strecke. Unter (möglichen) **Gewichten** eines gewichteten Mittelwerts $\overline{z}$ von u Zahlenwerten (oder Werten eines dimensionsbehafteten Skalars) z_i versteht man u positive Zahlen $g_i < 1$, deren Summe 1 ist (zum Beispiel die Wahrscheinlichkeiten, mit denen diese Werte in einem Feld zufällig erscheinen). Der **gewichtete Mittelwert** irgendwelcher u Werte z_i ist die Summe aller u Produkte $g_i z_i$ (im Beispiel also der Erwartungswert der Zufallsgröße z). Zusammenfassung:

2.5 *Bazaj teoremoj de la informaciteorio*

2.50 *Lemo: La malegalaĵo de Jensen.*

Pluraj bazaj teoremoj de la informaciteorio estas komforte pruveblaj helpe de la malegalaĵo de Jensen (bildo 2.50a). Ĉi tiu konstatas, ke ĉe funkcio $f(x)$ konveksa (en intervalo) la pezita mezvaloro $\overline{f}(x)$ estas senescepte malpli granda ol la funkcivaloro $f(\overline{x})$ de la pezita mezvaloro $\overline{x}$ de la koncernaj argumentvaloroj x_i, se ĉi tiuj ne ĉiuj koincidas. En ĉi tiu konteksto funkcio $f(x)$ nomitas **konveksa** (pli precize: „de supre“ konveksa) en la intervalo $]a, b[$, se - imagige parolante – por ajnaj du diversaj punktoj x_1, x_2 el ĉi tiu intervalo la streko inter la kurbopunktoj $f(x_1)$ kaj $f(x_2)$ situas komplete sub la kurbo bildiganta $f(x)$ – kun la escepto, kompreneble, de la komenca kaj la fina punktoj de ĉi tiu streko. Per (eblaj) **pezoj** de pezita mezvaloro $\overline{z}$ de u valoroj z_i de nombro (aŭ de dimensihava skalaro) oni komprenas u pozitivajn nombrojn $g_i < 1$, kiuj sumiĝas al 1 (ekz. la aperprobabloj de ĉi tiuj valoroj en kampo). La **pezita mezvaloro** (aŭ: pezita aritmo) de ajnaj u valoroj z_i estas la sumo de ĉiuj u produtoj $g_i z_i$ (en la ekzemplo do la ekspekto de la stokasta variablo z). Resumo:

(2.50.1) $$\sum_{i=1}^{u} g_i = 1$$

(2.50.2) $$\overline{z} := \sum_{i=1}^{u} g_i \cdot z_i$$

(2.50.3) $$\overline{f}(x) < f(\overline{x})$$

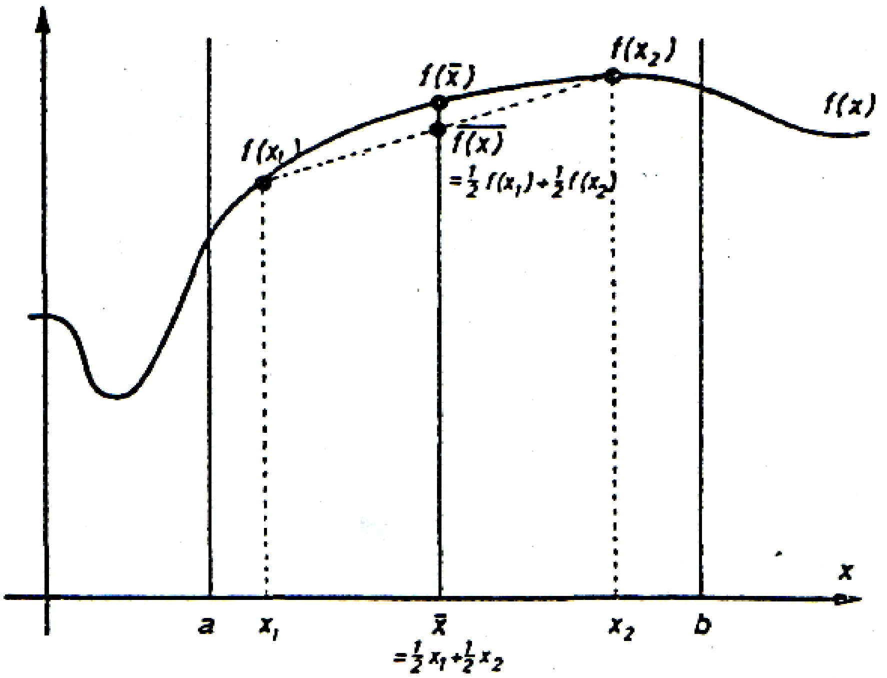

Bild 2.50a: *Die Jensensche Ungleichung im Spezialfall von nur u = 2 verschiedenen Argumentwerten gleichen Gewichts ($g_1 = g_2 = 1/2$)*

Bild 2.50a enthält das Schaubild einer Funktion, welche im Intervall]*a, b*[konvex ist[23]. Der gewichtete Mittelwert der u = 2 Werte x_1, und x_2 mit den Gewichten $g_1 = g_2 = 1/2$ ist natürlich der einfache arithmetische Mittelwert $\bar{x}$. Der Leser macht sich (z. B. unter Benutzung des Strahlensatzes) leicht klar. daß in diesem Spezialfall der gewichtete Mittelwert $\overline{f}(x)$ *der Funktionswerte* $f(x_1)$ und $f(x_2)$ gerade der Funktionswert der *Sehnenfunktion* an der Stelle $\bar{x}$ ist; er ist also sicher kleiner als der Funktionswert von $f(x)$ an derselben Stelle x. Dieser Satz läßt sich verallgemeinern auf *unterschiedliche* Gewichtszahlen und (z. B. durch vollständige Induktion) auf eine *beliebige* Anzahl $u \geq 2$ von Argumentwerten. In dieser Verallgemeinerung lautet ein Hilfssatz,

Bildo 2.50a: *La malegalaĵo de Jensen en la speciala kazo de nur u = 2 diversaj argumentvaloroj sampezaj ($g_1 = g_2 = ½$)*

La bildo 250a bildigas funkcion, kiu estas kompleksa en intervalo]*a, b*[[23]. La pezita mezvaloro de la u = 2 valoroj x_1 kaj x_2 kun la pezoj $g_1 = g_2 = 1/2$ estas kompreneble la simpla aritmo $\bar{x}$. La leganto evidentigas al si facile (ekz. uzante la raditeoremon), ke en ĉi tiu speciala kazo la pezita mezvaloro $\overline{f}(x)$ *de la funkcivaloroj* $f(x_1)$ kaj $f(x_2)$ estas ĝuste la funkcivaloro de la funkcio de la kordo je la punkto $\bar{x}$; ĝi do certe malsuperas la samlokan funkcivaloron de $f(x)$. Ĉi tiun teoremon eblas ĝeneraligi al *malsamaj* pezoj kaj (ekz. per matematika indukto) al *ajna* nombro $u \geq 2$ da argumentvaloroj. En ĉi tiu ĝeneraligo tekstas lemo, kiun ni uzos en la sekvantaj

[23]Diese Funktion ist ersichtlich sogar im links und rechts *geschlossenen* Intervall [*a, b*] konvex, d. h. es könnte z. B. auch $x_1 = a$ sein. Es ist aber für unsere Zwecke manchmal einfacher, vom beiderseitig *offenen* Intervall]*a, b*[auszugehen, zu welchem die Grenzpunkte *a* und *b* gerade nicht mehr gehören, so daß dort auch eine Unstetigkeit bestehen kann.

[23] Ĉi tiu funkcio evidente estas konveksa eĉ en la maldekstre kaj dekstre *fermita* intervalo [*a. b*], tiel ke povus validi ankaŭ $x_1 = a$. Sed por niaj celoj estas foje pli simple, argumenti per ambaŭflanke *malfermita* intervalo]*a, b*[, al kiu ĝuste la limpunktoj *a* kaj *b* ne plu apartenas, ebligante tiean malkontinuecon.

auf den wir in folgenden Kapiteln mehrere Beweise stützen werden: **Wenn**
(a) $f(x)$ in einem Intervall $]a, b[$ konvex ist,
(b) $x_i \in]a, b[$ für alle $i \leq u$, wobei $u \geq 2$,
(c) $x_i \neq x_j$ für mindestens ein Indexpaar i, j (falls also nicht alle Argumentwerte zusammenfallen) und
(d) für $u \geq 2$ positive Zahlen g_i die Gleichung (2.50.1) gilt,
***dann** gilt die Jensensche Ungleichung:*

ĉapitroj por pluraj pruvoj, jene: **Se**
(a) $f(x)$ estas konveksa en intervalo $]a, b[$,
(b) $x_i \in]a, b[$ por ĉiuj $i \leq u$, kun $u \geq 2$,
(c) $x_i \neq x_j$ por almenaŭ unu paro da indicoj i, j (se do ne ĉiuj argumentvaloroj koincidas) kaj
(d) por $u \geq 2$ pozitivaj nombroj g_i validas la egalaĵo (2.50.1),
***tiam** validas la malegalaĵo de Jensen:*

(2.50.4) $$\bar{f}(x) := \sum_{i=1}^{u} g_i \cdot f(x_i) < f(\sum_{i=1}^{u} g_i \cdot x_i) =: f(\bar{x})$$

2.51 *Maximale Unsicherheit H, Knappheit und Redundanz.*

2.51 *Maksimuma necerteco H, koncizeco kaj redund*

In Kapitel 2.3 wurde der Selbstbeobachtung entnommen und an einigen Beispielen verdeutlicht, dass unser Gefühl der Unsicherheit bei gleichem Feldumfang desto schwächer wird, je stärker die einzelnen Wahrscheinlichkeiten voneinander abweichen. Mit der Jensenschen Ungleichung können wir nun umgekehrt mathematisch beweisen, dass - bei konstantem Feldumfang $u(Z)$ - das Unsicherheitsmaß H sein Maximum im Falle der Gleichwahrscheinlichkeit annimmt. Für $u(Z) = 2$ (Bild 2.3a) wurde dies in Kapitel 2.3 schon auf schulmathematische Weise bewiesen. Zu beweisen ist noch der allgemeine

Satz: Die Unsicherheit H in einem Feld Z, dessen $u = u(Z) \geq 2$ Elemente nicht alle die gleiche Wahrscheinlichkeit[24] haben, ist

En la ĉapitro 2.3 estis ekkonate el la memobservado kaj ilustrite per kelkaj ekzemploj, ke nia sento de malcerteco kaze de egalampleksaj kampoj des pli malfortiĝas, ju pli diferencas inter si la unuopaj probabloj. Per la malegalaĵo de Jensen ni nun povas inverse matematike pruvi, ke – kaze de konstanta amplekso $u(Z)$ – la mezuro H de la malcerteco atingas sian maksimumon kaze de egalprobableco. Por $u(Z) = 2$ (bildo 2.3a) tio estis jam pruvata laŭ lernejmatematika maniero en ĉapitro 2.3. Pruvenda estas ankoraŭ la ĝenerala

Teoremo: La malcerteco H en kampo Z, kies $u = u(Z) \geq 2$ elementoj ne ĉiuj samprobablas[24], estas

(2.51.1) $$H/\text{bit}^{25} < \text{ld } u$$

[24] Hier und bei allen folgenden Anwendungen der Jensenschen Ungleichung kann man alle Elemente der Wahrscheinlichkeit 0 aus dem Feld zuvor schon streichen. Das setzen wir beim Beweis als geschehen voraus. Der Satz gilt aber *a fortiori* auch für das ursprüngliche, umfangreichere Feld!

[24] Ĉi tie kaj ĉe ĉiuj sekvaj aplikoj de la malegalaĵo de Jensen eblas jam anticipe forstreki el la kampo ĉiujn elementojn kun probablo 0. Ni supozas dum la pruvo ke tio jam okazis. Des pli la teoremo validas por la origina, pli ampleksa kampo!

Sie ist kleiner als die im Falle der Gleichwahrscheinlichkeit (d.h. für $p_i \equiv 1/u$) bestehende maximale Unsicherheit

Ĝi estas malpli granda ol la maksimuma necerteco reganta kaze de egalprobableco (t.e. por $p_i \equiv 1/u$)

(2.51.2) $$H_{\max}/\text{bit} = \text{ld}\, u$$

Beweis:

Wegen $p_i > 0$ und $u \geq 2$ liegen die Wahrscheinlichkeiten p_i alle in]0; 1[, erfüllen also die Bedingung (d) für Gewichtszahlen in der Jensensche Ungleichung: $g_i \equiv p_i$. Ferner ist die Funktion *ld x* im Intervall]1; ∞[konvex, erfüllt dort also die Bedingung (a) für eine Funktion in der Jensenschen Ungleichung: $f(x) \equiv \text{ld}\, x$. Schließlich sind die Unwahrscheinlichkeiten $1/p_i$ nicht alle gleich, und sie liegen im Konvexitätsintervall; sie können daher gemäß (*b, c*) als Argumentwerte x_i in die Jensensche Ungleichung eingesetzt werden: $x_i \equiv 1/p_i$. Mit diesen Einsetzungen wird aus (2.50.4) speziell:

Pruvo:

Pro $p_i > 0$ kaj $u \geq 2$ la probabloj p_i situas ĉiuj en]0; 1[. Ili do plenumas la kondiĉon (d) por roli kiel pezoj en la malegalaĵo de Jensen: $g_i \equiv p_i$. Krome la funkcio ld x estas konveksa en]1; ∞[. Ĝi do tie plenumas la kondiĉon (a) por roli kiel funkcio en la malegalaĵo de Jenssen: $f(x) \equiv \text{ld}\, x$. Fine la malprobabloj $1/p_i$ ne ĉiuj egalas, kaj ili situas en la konveksecintervalo; do laŭ (*b, c*) eblas ilin enmeti kiel argumentvalorojn en la malegalaĵon de Jensen: $x_i \equiv 1/p_i$. Per ĉi tiuj substitucioj (2.50.4) specialiĝas al:

(2.51.3) $$\sum_{i=1}^{u} p_i \cdot \text{ld}\frac{1}{p_i} < \text{ld}\Big(\sum_{i=1}^{u} p_i \cdot \frac{1}{p_i}\Big) = \text{ld}\Big(\sum_{i=1}^{u} 1\Big) = \text{ld}\, u$$

Die linke Seite dieser Ungleichung ist nach Gleichung (2.19.3) die Unsicherheit H/bit. Auf der rechten Seite ist die Summe von u Summanden zu bilden, die alle gleich 1 sind; diese Summe ist also gleich u, so daß rechts *ld u* steht. Damit ist (2.51.1) für alle $u \geq 2$ bewiesen.

Im Falle der Gleichwahrscheinlichkeit haben alle u Möglichkeiten des Feldes dieselbe Wahrscheinlichkeit, nämlich $1/u$, und damit dieselbe Information, nämlich ld u. Setzt man diese speziellen Werte in Gleichung (2.19.3) ein, dann erhält man u gleiche Summanden, folglich die Summe ld u als Unsicherheit bei dieser bisher ausgenommenen Wahrscheinlichkeitsverteilung.

La maldekstra flanko de ĉi tiu malegalaĵo estas laŭ (2.19.3) la necerteco H/bit. Je la dekstra flanko adiciendas u adiciatoj, kiuj ĉiuj egalas al 1; ĉi tiu sumo do estas u, tiel ke dekstre staras ld u. (2.51.1) tiel estas pruvita por ĉiuj $u \geq 2$.

En la kazo de egalprobableco ĉiuj u ebloj de la kampo havas la saman probablon, nome $1/u$, sekve la saman informacion, nome ld u. Se oni enmetas ĉi tiun specialan valoron en la egalaĵon (2.19.3), ĉiuj adiciatoj egalas. Kiel necertecon en ĉi tiu ĝis nun ekskludita probablodistribuo oni do ricevas la sumon ld u. Pro tio, ke krome ĉi valo-

[25] Statt der strengen Schreibweise „H/bit < ld u" - ld u ist eine bloße Zahl, während H nach unserer Konvention die Pseudomaßeinheit „bit" tragt - ist es meist blich, einfacher „H < ld u" zu schreiben, oder - besser - „H < log u".

[25] Anstataŭ rigore skribi „H/bit < ld u" – ld u estas nura nombro, sed H havas laŭ nia konvencio la pseŭdomezurunuon „bit" – oni kutimas skribi pli simple „H < ld u" - aŭ, pli bone - „H < log u".

Da dieser Wert von H im übrigen - nach der für alle anderen Verteilungen schon bewiesenen Ungleichung (2.51.1) - größer ist als H für die anderen Verteilungen, ist er das Maximum von H. Damit ist auch (2.51.2) bewiesen.

Durch Kombination der Beziehungen (2.51.1) und (2.51.2) erhält man die (weniger aussagekräftige) allgemeine Beziehung

ro de H - pro la malegalaĵo (2.51.1), pruvita jam por ĉiuj aliaj distribuoj – estas pli granda ol H por la aliaj distribuoj, ĝi estas la maksimumo de H. Tiel pruviĝis ankaŭ (2.51.2).

Per kombino de la rilatoj (2.51.1) kaj (2.51.2) oni ricevas la (malpli informigan) ĝeneralan rilaton

$$H/\text{bit} \leq \text{ld } u \tag{2.51.3}$$

Eine Nachrichtenquelle möge wiederholt Zeichen aus dem endlichen, konstanten Feld Z (Umfang: $u(Z) =: u$) liefern. Die jeweilige Information ist dabei eine Zufallsgröße. Ihr Erwartungswert ist nach Gleichung (2.19.3) gerade $H(Z) =: H$. Es ist, wie bei jeder Zufallsgröße, nach dem in Abschnitt 2.19 zu den Gleichung (2.19.1) und (2.19.2) Gesagten zu erwarten, daß die *durchschnittliche* Information der Zeichen bei wachsender Länge N der Folge gegen den Erwartungswert strebt, also im Falle der Zufallsgröße „Information des gesendeten Zeichens" gegen H, daß also die *Folge insgesamt* eine Information der Größenordnung $N{*}H$ enthält.

Ist diese Information kleiner als $N{*}\text{ld } u$ bit, dann heißt die Nachrichtenquelle bzw. die von ihr gelieferte Zeichenfolge „redundant", deutsch: „weitschweifig".

Diese Bezeichnung ist treffend, denn mit N Zeichen hätte die Quelle nach (2.51.2) die Information $N{*}\text{ld } u$ bit liefern können, hätte sie nur alle Zeichen gleichwahrscheinlich gesendet. Sie hat also, umgangssprachlich formuliert, „zu wenig mit zuviel Worten" gesagt (wenn das Zeichenrepertoire aus Wörtern besteht). Anders ausgedrückt: eine *nicht* redundante Quelle hätte zur Lieferung der Information NH *weniger* als N Zeichen gebraucht, nämlich N^*, die aber alle $H_{max} = \text{ld } u$ bit Information enthalten hätten.

Es ist also $N^*{*}\text{ld } u \textit{ bit} = N{*}H$.

Der Bruchteil, auf welchen die Zeichen-

Mesaĝfonto havigu ripete signojn el la finia, konstanta kampo Z (amplekso: $u(Z) =: u$). Ilia informacio en la unuopa kazo estas stokasta valoro. Ĝia ekspektvaloro estas laŭ egalaĵo (2.19.3) ĝuste $H(Z) =: H$. Kiel ĉe ĉiu stokasta valoro laŭ la konstataĵoj faritaj pri (2.19.1) kaj (2.19.2) en ĉapitreto 2.18 estas ekspekteble, ke la *aritma* infomarcio de la signoj por kreskanta longeco N de la sinsekvo konverĝas al la ekspektvaloro, do kaze de la stokasta variablo „informacio de la sendita signo" al H, do ke la *sinsekvo entute* enhavas proksimume la informacion $N{*}H$.

Se ĉi tiu informacio malsuperas $N{*}\text{ld } u$ bit, tiam la informfonto resp. la de ĝi sendita signovico nomitas „redunda" (malkonciza).

Ĉi tiu nomo estas trafa, ĉar per N signoj la fonto povintus havigi laŭ (2.51.2) la informacion $N{*}\text{ ld } u$ bit, se ĝi nur estus sendinta ĉiujn signojn samprobable. Per ĉiutaga parolturno esprimite: la fonto diris „tro malmulton per tro multaj vortoj" (se la signorepertuaro konsistas el vortoj). Alie esprimite: *ne* redunta fonto povintus liveri la informacion NH per *malpli* ol N signoj, nome per N^*, de kiuj tamen ĉiuj enhavintus $H_{maks} = \text{ld } u$ bit da informacio.

Do $N^*{*}\text{ld } u \textit{ bit} = N{*}H$.

La ono, al kiu signovico ŝrumpus,

folge zusammenschrumpfen würde, wenn auf Weitschweifigkeit verzichtet würde, ist nach dieser Gleichung

se estus rezignite pri malkoncizeco, laŭ ĉi tiu egalaĵo estas

$$k := \frac{N^*}{N} = \frac{H\ /\ \mathrm{bit}}{\mathrm{ld}\ u} \qquad (2.51.5)$$

Dieser Bruchteil heißt zurecht „Knappheit“. Ist nämlich für eine Zeichenfolge k *klein*, dann entält schon eine erheblich kürzere andere, redundanzfreie Zeichenfolge gleich viel Information[26] wie die betrachtete, so dass diese wenig knapp, vielmehr sehr weitschweifig (redundant) ist. Die höchste Knappheit $k = 1$ wird erreicht, wenn $N^* = N$ ist, also gleichviel Information nicht mit einer kürzeren Folge lieferbar ist.

Ĉi tiu ono nomitas trafe „koncizeco“. Ĉar se k de signovico *malgrandas*, tiam malpli longa, alia, neredunda signovico enhavus la saman kvanton da informacio[26] kiel la konsiderata, tiel ke ĉi tiu ne tre koncizas sed malkoncizas (redundas). La maksimuma koncizeco $k = 1$ estas atingita, se $N^* = N$, se do sammulton da informacio ne havigeblas per malpli longa vico.

Der Bruchteil oder Prozentsatz, der bei knappest möglicher Informationpackung eingespart werden könnte, heißt „relative Redundanz".

La ono aŭ procentaĵo, kiun eblus ŝpari, se oni plejeble koncizigas la ensignigadon de la informacio, nomitas „relativa redundo“.

$$r := 1 - k = 1 - \frac{H\ /\ \mathrm{bit}}{\mathrm{ld}\ u} = \frac{\mathrm{ld}\ u - H\ /\ \mathrm{bit}}{\mathrm{ld}\ u} \qquad (2.51.6)$$

Als „absolute Redundanz" wird der Ausdruck

„Absoluta redundo“ nomitas la esprimo

$$R := \mathrm{ld}\ u\ \mathrm{bit} - H \qquad (2.51.7)$$

bezeichnet, also der Fehlbetrag der Unsicherheit gegenüber der maximalen Unsicherheit.

do la manko de la necerteco ĝis la maksimuma necerteco.

In den beiden nächsten Kapiteln werden wir erkennen, daß eine stochastische Abhängigkeit in der Aufeinanderfolge der Zeichen einer Folge deren Redundanz stets vergrößert.

En la du sekvantaj ĉapitroj ni ekkonos, ke stokastika dependeco, laŭ kiu la signoj sin sekvas, ĉiam plialtigas la redundon de la sinsekvo.

[26] Damit ist *noch nicht* gesagt, dass es sich um die selbe *Information (Nachricht)* handeln könnte. Man kann ja auch die mit einem Aluminiumwürfel gelieferte Masse mit einem wesentlich kleineren Goldwürfel liefern (weil die spezifische Dichte von Gold größer ist), aber die Materie der beiden Liefe-rungen ist verschieden. Da aber (im Unterschied zur Materie) ein Zeichen nicht *„ist“* sondern *„zeigt“,* enthält die redundanzfreie und daher kür-zere Zeichenfolge tatsächlich nicht nur (quanti-ta-tiv) *gleich viel* Information wie die längere, sondern kann (bei passender Umkodierung) auch *qualitativ die selbe* Information liefern. Dies wird in Teil III bewiesen.

[26] Per tio *ankoraŭ ne* estas dirite, ke povus temi pri la *sama informo*. Oni ja povas an-kaŭ la mason, liveritan per kubo el alumi-nio, liveri per konsiderinde malpli granda kubo el oro (ĉar la specifa denseco de oro estas pli granda), sed la materio de la du liveroj malegalas. Sed pro tio, ke (alie ol materio) signo ne *„estas“* sed *„montras“*, la ne redunda kaj tial malpli longa signosin-sekvo fakte enhavas ne nur (laŭkvante) *sammulton* da informacio kiel la pli longa, sed povas (se ĝi estas taŭge alikodita) liveri ankaŭ *laŭkvalite la saman* informon. Tio estos pruvata en parto III.

2.52 *Superierungsredundanz*

Dieselbe, von einer Quelle gesendete Signalfolge kann, wie gegen Ende des Kapitels 2.18 erläutert, den Empfang unterschiedlicher Zeichenfolgen bewirken, je nachdem, wie der Empfänger gerade superiert, also in Abhängigkeit von seinem Zustand. Man kann z.B. Buchstabenfolgen oder (komplexbildend) Wortfolgen lesen. Beim Lesen einer in unterschiedlichen Schrifttypen gedruckten Folge von Buchstaben kann man von den Typen (klassenbildend) abstrahieren. Zwei Empfänger können daher dieselbe Signalfolge auf zwei verschiedenen Stufen einer Superzeichenhierarchie apperzipieren. Durch klassenbildendes Superieren verliert eine Folge an „unwesentlicher“ Information. (Man superiert klassenbildend ja gerade durch Abstraktion von dem zum jeweiligen Zwecke Unwesentlichen!). Auch bei komplexbildendem Superieren kann man dies vermuten.

In Anlehnung[27] an eine Begriffsbildung von Felix von Cube (1965, S. 115) bezeichnen wir als „*Superierungsredundanz*" den Ausdruck

2.52 *Prosuperumadredundo*

La sama, de fonto sendita signalsinsekvo povas, kiel klarigite fine de la ĉapitro 2.18, kaŭzi la ricevon de diversaj signo-sinsekvoj, depende de tio, kiel la ricevanto aktuale superumas, do depende de lia stato. Oni povas legi ekz. litersinsekvojn aŭ (kompleksige) vortsinsekvojn. Legante litersinsekvon presitan per diversaj litertipoj one povas de la tipoj (klasige) abstrakti. Du diversaj ricevantoj do povas apercepti la saman signalsinsekvon sur du diversaj ŝtupoj de hierarkio de „supersignoj“ (= kunsignoj). Per klasiga superumo la sinsekvo perdas „ne esencan“ informon. (Oni ja superumas klasige per abstrakto de tio, kio ne estas esenca por la aktuala celo!) Oni povas konjekti tion ankaŭ ĉe la kompleksiga superumado.

Apogante[27] nin al nocikreaĵo de Felix von Cube (1965, p 115) ni no mas „*prosuperumadredundo*“ la esprimon

$$r_s := 1 - I_s/I \qquad (2.52.1)$$

in welchem I die Gesamtinformtion einer Zeichenfolge (z.B. Buchstabenfolge) ist, I_s die Gesamtinformation derselben gesendeten Nachricht, die aber der Empfänger als Superzeichenfolge (z.B. als Folge ganzer Wörter) „apperzipiert“ (d.h. bewusst wahrnimmt).

en kiu I signifas la tutan informacion de signosinsekvo (ekz. de litersinsekvo), I_s la tutan informacion de la sama, sendita mesaĝo, sed „apercep-tita“ (t.e. enkonsciigita) de la ricevonto kiel sinsekvon de supersignoj (ekz. de vortoj).

[27] Das durch F. von Cube eingeführte Maß der „Superzeichenredundanz“ weicht von der Definitionsgleichung (2.52.1) der Superierungsredundanz etwas ab. Erstens führt es F. von Cube nur für den Fall komplexbildender Superierung ein, während (2.52.1) auch bei Klassenbildung gilt. Zweitens stimmt v.Cubes Maß auch in diesem Spezialfall nur näherungsweise mit dem hier definierten Maß r_s überein. Mit diesem dürfte aber v.Cubes Grundgedanke verallgemeinert und präzisiert sein.

[27] La mezuro de „supersigna redundo“ enkondukita de F. von Cube iom devias de la prosuperumadredundo difinita per (2.52.1). Unue F. von Cube ĝin enkondukas nur por la kazo de kompleksiga superumado, dum kiam (2.52.1) validas ankaŭ por klasigo. Due ankaŭ en ĉi tiu spe-ciala kazo la mezuro de v.Cube nur proksimume koincidas kun la ĉi tie difinita mezuro r_s. Sed ĉi tiu supozeble ĝeneraligas kaj plipreciziĝas la bazan ideon de v.Cube.

Die Fruchtbarkeit dieser Begriffsbildung besteht darin, dass sie erstmals einen Zusammenhang zwischen bloß statistischen Größen und der Kategorie des *Nutzens* herstellt Die *Richtung* der Superierung - und damit die Superzeichenredundanz! - hängt ja vom *Zweck* der lnformationsaufnahme ab.

Die Superierungsredundanz im eben definierten Sinne darf nicht verwechselt werden mit der Redundanz der auf einer Superzeichenstufe apperzipierten Folge. Beispielsweise möge eine Quelle mit gleicher Wahrscheinlichkeit und unabhängig vom jeweiligen Vorgänger die vier Fragewörter und die entsprechenden vier Zeigewörter von ILo für Individuen (KIU, TIU), Klassen (KIO, TIO), Eigenschaften (KIA, TIA) und Orte (KIE, TIE) liefern, so dass z.B. in einer N_1 = 80 Textwörter langen Folge jedes dieser 8 Wortschatzwörter 10 Mal auftritt. Der Informationsgehalt der Folge beträgt also $I_1=80 \cdot \mathrm{ld}\ 8$ bit=240 bit. Interessiert man sich nur dafür, ob das jeweils nächste Wort fragt oder zeigt, superiert man also klassenbildend zu den nur zwei Superzeichen {K..} und {T...}, dann apperzipiert man auf dieser Superzeichenstufe eine gleichlange Zeichenfolge, jedoch mit dem auf $I_2=80 \cdot \mathrm{ld}\ 2$ bit = 80 bit gesunkenen Informationsgehalt. Nach der Definitionsgleichung (2.52.1) berechnet man als Superierungsredundanz r_s = 2/3. Da aber die u = 2 verschiedenen Superzeichen mit gleicher Wahrscheinlichkeit aufeinanderfolgen, ist die Unsicherheit auf dieser Betrachtungsstufe H =1 bit, also die Redundanz nach Definitionsgleichung (2.51.6) r = 0 (wie auch bei *Beachtung der Verschiedenheit* der u = 8 Wörter auf der darunter liegenden Stufe, auf welcher die Unsicherheit über das jeweils nächste Wort H = ld 8 bit = 3 bit beträgt).

Auf die Stufe der *Wörter* gelangt man komplexbildend von der Stufe der *Schriftzeichen* aus: jedes der 8 Wörter besteht aus 3

La starigo de ĉi tiu nocio estas fruktodona, ĉar ĝi la unuan fojon interrilatigas nure statistikajn valorojn kun la kategorio de la *utileco.* La *direkto* de la superumado - kaj per ĝi la prosuperumada redundo! - ja dependas de la *celo* de la informenpreno.

La prosuperumadredundo en la ĵus difinita senco ne estu konfuzata kun la redundo de sinsekvo aperceptata sur ŝtupo de supersignoj. Ekzemple liveru fonto samprobable kaj stokaste sendepende la kvar ILajn demandvortojn kaj la korespondajn kvar montro-vortojn por individuoj (KIU, TIU), klasoj (KIO, TIO), kvalitoj (KIA, TIA) kaj lokoj (KIE, TIE), tiel ke en ekz. N_1 = 80 tekstvortojn longa sinsekvo ĉi tiuj 8 vorttrezoraj vortoj aperas po 10-foje. La informacio de la sinsekvo do estas $I_1=80 \cdot \mathrm{ld}\ 8$ bit = =240 bit. Se oni interesiĝadas nur, ĉu la sekvanta vorto demandas aŭ montras, se oni do superumadas klasige al la nur du supersignoj {K...} kaj {T...}, tiam oni aperceptadas sur ĉi tiu supersigna ŝtupo samlongan signosinsekvon, sed kun informacio malpligrandigita al I_2 = $80 \cdot \mathrm{ld}\ 2$ bit = =80 bit. Laŭ la difina egalaĵo (2.52.1) oni trovas kiel prosuperumadredundanco r_s = 2/3. Sed ĉar la u = 2 diversaj supersignoj sinsekvas samprobable, la necerteco sur ĉi tiu aperceptadŝtupo estas H = 1 bit, do la redundo laŭ la difinegalaĵo (2.51.6) r = 0 (same kiel kaze de *atentado de la diverseco* de la u = 8 vortoj sur la senpere suba ŝtupo, sur kiu la necerteco pri la sekvanta vorto estas H = ld 8 bit = 3 bit).

La ŝtupon de la *vortoj* oni atingas kompleksige de la ŝtupo de la *skribsignoj*: ĉiu de la 8 vortoj konsistas el 3

Buchstaben, gefolgt von einem Zwischenraum. Statt N_1=80 Wortzeichen ist also dieselbe, von der Quelle gelieferte Nachricht N_0=320 Schriftzeichen lang. Von ihnen sind I und der Wortzwischenraum die häufigsten: ihre Auftrittswahrscheinlichkeit ist je (80/320 =)¼. K und T treten je mit der Wahrscheinlichkeit 1/8 auf, und die vier Schlussvokale an je 20 Stellen der Schriftzeichenfolge - sie haben je die Auftrittswahrscheinlichkeit 1/16. Durch Addition der Informationsbeiträge oder durch Anwendung der Definitionsgleichung (1.2.3) berechnet man als Informationsgehalt der 320 Schriftzeichen $I_0=N_0H_0=320 \cdot 2{,}75$ bit=880 bit. Ihre Redundanz ist 1–2,75/3=1/12= =0,083..., ihre Knappheit 11/12. Kehrt man zurück zur Superzeichenstufe der Wortzeichen, also zu den nur 240 bit, dann „erlebt" man eine Superierungsredundanz von 1–240/880=8/11=0,7272... (eine Superierungsverknappung auf 3/11). Um rund 73% (auf 3/11) sinkt also die Information (die Unvorhersehbarkeit), sobald man die Nachricht nicht mehr als stochastisch unabhängige Folge von Schreibmaschinenanschlägen sondern als stochastisch unabhängige Wortfolge ansieht (d.h. so apperzipiert).

Dass dieser Wert erheblich größer ist als die berechnete Redundanz der Schriftzeichenfolge (nur 8,3... %), beruht offensichtlich auf der starken Abhängigkeit, mit welcher diese Zeichen aufeinanderfolgen: auf K und T folgt *immer* I, auf U, O, A und E folgt *immer* der Zwischenraum. Diese beiden Zeichen haben also hier die Wahrscheinlichkeit 1 und damit die Information 0. Entsprechend ist die bedingte Wahrscheinlichkeit von U, O, A und E nach I je ¼ (bedingte Information: 2 bit), von K und T nach dem Zwischenraum je ½ (1 bit Information). Zählt man die insgesamt 320 Informationswerte der je 80 Zwischenräume und Buchstaben I (0 bit), der je 40 Zei-

literoj, al kiu sekvas interspaco. Anstataŭ N_1=80 vortsignoj la sama, de la fonto liverata mesaĝo do longas N_0=320 skribsignojn. De ili I kaj la interspaco estas la plej oftaj: ilia aperprobablo estas po (80/320=)¼. K kaj T aperas kun probablo po 1/8, kaj la kvar vortfinigaj vokaloj aperas ĉe po 20 pozicioj de la sinsekvo de skribsignoj – ilia aperprobablo estas po 1/16. Adiciante la informacikontribuojn aŭ aplikante la difinegalaĵon (1.2.3) oni elkalkulas kiel informacion de la 320 skribsignoj $I_0=N_0H_0=320 \cdot 2{,}75$ bit = 880 bit. Ilia redundo estas 1 – 2,75/3 = 1/12 = 0,0833..., ilia koncizeco 11/12. Se oni revenas al la supersigna ŝtupo de la vortsignoj, do al la nur 240 bit, tiam oni „sentas" prosuperumadredundon de 1 – 240/880 = =8/11 = 0,7272... (prosuperumadkoncizigon al3/11). Je ĉirkaŭ 73% (al 3/11) do malkreskas la informacio (la neantaŭvidebleco), kiam oni ne plu konsideras (t.e. aperceptas) la mesaĝon kiel stokastike sendependan sinsekvon de skribmaŝinaj signoj, sed kiel tian vortsinsekvon.

Ke ĉi tiu valoro estas multe pli granda ol la elkalkulita redundo de la sinsekvo de skribsignoj (nur 8,3...%), estas evidente sekvo de la forta dependeco en la signosinsekvo: al K kaj T *ĉiam* sekvas I, al U, O, A kaj E *ĉiam* sekvas interspaco. Ĉi tiuj du signoj do havas ĉiloke la probablon 1, sekve la informacion 0. Analoge la kondiĉita probablo de U, O, A kaj E post I estas po ¼ (kondiĉita informacio: 2 bit), de K kaj T post interspaco po ½ (1 bit da informacio). Se oni adicias la entute 320 informacivalorojn de la po 80 interspacoj kaj literoj I (0 bit), de la po 40 signoj K kaj T (po 1 bit) kaj de la chen K und T (je 1 bit) und der je 20 Zeichen U, O, A und E (je 2 bit) zusammen, dann erhält man nur noch die 240 bit der Wortfolge. Verteilt auf die 320 Schriftzeichen ergibt dies die (mittlere bedingte) Unsicherheit $H_X(Y)$= 0,75 bit pro Schriftzeichen statt des Maximums ld 8=3 bit/Schriftzeichen. Die Redundanz beträgft also 75%. (die Knappheit k = ¼).

Die Superierungsredundanz ist bei Komplexbildung eine Auswirkung der stochastischen Abhängigkeit zwischen den Zeichen eines Komplexes. Denn diese senkt, wie anschließend bewiesen wird, im Mittel deren Informationsgsgehalt.

po 20 signoj U, O, A kaj E (po 2 bit), tiam oni trovas nur la 240 bit de la vortsinsekvo. Distribuite al la 320 skribsignoj rezultas la (aritma kondiĉita) necerteco $H_X(Y)$ = 0,75 bit je skribsigno anstataŭ la maksimumo ld 8 = 3 bit/skribsigno. La redundo do estas 75% (la koncizeco k = ¼).

La prosuperumadredundo estas kaze de kompleksigo sekvo de la stokasta dependeco inter la signoj de komplekso. Ĉar ĉi tiu malplialtigas, kiel sekve pruvate, aritme ties informacion.

2.53 *Redundanz durch stochastische Abhängigkeit*

2.53 *Redundo pro stokastika dependeco*

Hinter der in Abschnitt 2.52 im Spezialfall von Zeichenfolgen erkannten lnformationsverringerung durch stochastische Abhängigkeit steckt folgender allgemeine

Satz: Sind die Felder Y und Z voneinander stochastisch abhängig, dann gilt

La informaciŝrumpo pro stokasta dependeco, rimarkita en la ĉapitreto 2.52 en la speciale kaze de signovicoj, radikas en la ĝenerala

Teoremo: Por kampoj Y kaj Z stokaste dependaj validas

(2.53.1) $$(H_Y(Z) =)\ \ H(Z|Y) < H(Z).$$

Sind die Felder voneinander *un*abhängig, dann gilt stattdessen

Se la kampoj estas stokaste *sen*dependaj, tiam validas anstataŭe

(2.53.2) $$(H_Y(Z) =)\ \ H(Z|Y) = H(Z).$$

Allgemein gilt also die (weniger aussagekräftige) Ungleichung

Ĝenerale do validas la (malpli informiga) rilato

(2.53.3) $$(H_Y(Z) =)\ \ H(Z|Y) \leq H(Z).$$

Im Durchschnitt (nicht in jedem Einzelfall!) kann also die Unsicherheit über die Nachricht aus einem Feld Z nur kleiner werden oder gleichbleiben, sobald eine Nachricht aus einem anderen Felde Y vorliegt.

Beweis:

Da Y und Z voneinander stochastisch abhängig sind, gibt es nach Abschnitt 2.16 mindestens ein $y_i \in Y$ und ein $z_k \in$Z, so dass $p(z_k|y_i) \neq p(z_k)$. Es muss auch (mindestens) ein weiteres y_j geben, so dass $p(z_k|y_i) \neq p(z_k|y_j)$ ist. Wären nämlich

Mezume (ne en ĉiu unuopa kazo!) la necerteco pri mesaĝo el kampo Z nur povas malpligrandiĝi aŭ resti konstanta, kiam ekestas mesaĝo el alia kampo Y.

Pruvo:

Ĉar Y kaj Z estas stokastike dependaj, ekzistas laŭ ĉapitreto 2.16 almenaŭ unu $y_i \in Y$ kaj unu $z_k \in$Z, tiel ke $p(z_k|y_i) \neq p(z_k|y_j)$. Devas ekzisti ankaŭ (minimume) unu kroma y_j, tiel ke $p(z_k|y_i) \neq p(z_k)$. Ĉar se ĉiuj kondiĉitaj alle bedingten Wahrscheinlichkeiten von z_k gleich, dann hinge dessen Wahrscheinlichkeit nicht von der Nachricht aus Y ab.

Die zu beweisende Ungleichung folgt nun aus einer Spezialisierung der Jensenschen Ungleichung (2.50.4). Man erhält sie, wenn man dort statt der allgemeinen Funktion $f(x)$ speziell die Funktion $x{*}\mathrm{ld}\,1/x$ einsetzt, die - wie Bild 2.24a zeigt - im Intervall]0,1[ja konvex ist, also die Bedingung (a) der Jensenschen Ungleichung erfüllt. Als Argumentwerte x_i setzen wir die $p(z_k|y_i)$ ein, und zwar für jenes z_k, für welches es mindestens ein y_i und ein y_j mit $p(z_k|\, y_i) \neq p(z_k)$ und $p(z_k|\, y_i) \neq p(z_k|\, y_j)$ gibt. Für diese Argumentwerte gelten daher die Jensenschen Bedingungen (b) und (c). Als Gewichte g_i, welche der Bedingung (d) genügen, können wir die $p(y_i)$ wählen. Aus (2.50.4) wird damit

probabloj de z_k estus egalaj, tiam ĝia probablo ne dependus de la mesaĝo el Y.

La pruvenda malegalaĵo sekvas el specialigo de la JENSENa malegalaĵo (2.50.4). Oni ĝin ricevas tie substituante la ĝeneralan funkcion $f(x)$ per la speciala funkcio $x{*}\mathrm{ld}\,1/x$, kiu ja estas (laŭ la bildo 2.24a) en]0; 1[konveksa, do plenumas la JENSENan kondiĉon (a). Kiel argumentvalorojn x_i ni enmetas la $p(z_k|y_i)$ por tiu z_k, por kiu ekzistas almenaŭ unu y_i kaj unu y_j kun $p(z_k|\, y_i) \neq p(z_k)$ kaj $p(z_k|\, y_i) \neq p(z_k|\, y_j)$. Por ĉi tiuj argumentvaloroj do validas la JENSENaj kondiĉoj (b) kaj (c). Kiel pezojn g_i plenumantajn (d) eblas elekti la $p(y_i)$. Tiel (2.50.4) fariĝas

$$\sum_i p(y_i)\cdot p(z_k\mid y_i)\cdot ld\,\frac{1}{p(z_k\mid y_i)} <$$

$$(2.53.4)\qquad < \Big(\sum_i p(y_i)\cdot p(z_k\mid y_i)\Big)\cdot ld\,\frac{1}{\sum_i p(y_i)\cdot p(z_k\mid y_i)}$$

Die rechte Seite dieser Ungleichung ist aber nach Gleichung (2.16.4) - nach Addition der Wahrscheinlichkeiten aller Kombinationen von z_k mit den verschiedenen y_i zur Wahrscheinlichkeit von z_k - identisch mit dem Ausdruck

La dekstra flanko de ĉi tiu malegalaĵo estas laŭ egalaĵo (2.16.4) – adiciinte la probablojn de ĉiuj kombinaĵoj de z_k kun la diversaj y_i al la probablo de z_k - identa kun la algebra esprimo

$$(2.53.5)\qquad \iota(z_k) = p(z_k)\cdot \mathrm{ld}\,\frac{1}{p(z_k)}$$

den wir in Kapitel 2.42 „informationelle Ausnutzung zugunsten von z_k" nannten. Diesen Ausdruck verwendend, können wir die schon bewiesene Ungleichung (2.53.4) kürzer schreiben:

kiun ni nomis en ĉapitro 2.42 „informeca ekspluato favore al z_k". Utiligante ĉi tiun esprimon ni povas plikoncizigi la jam pruvitan malegalaĵon (2.53.4)

$$(2.53.6)\qquad \Sigma\, p(y_j)\cdot \iota(z_k|\,y_j) < \iota(z_k)$$

Die Ungleichung (2.53.4) bzw. (2.53.6) gilt *nur* für jene z_k, für welche die $p(z_k|y_i)$ *nicht* durchweg gleich $p(z_k)$ sind. Für die übrigen z_k ist offensichtlich auch die linke Seite von (2.53.4) gleich der informationellen Ausnutzung und damit gleich der rechten Seite.

Summiert man daher links und rechts über alle k, dann bleibt das Kleinerzeichen gültig, und man erhält:

La malegalaĵo (2.53.4) resp. (2.53.6) *nur* validas por tiuj z_k, por kiuj la $p(z_k|y_i)$ *ne* senescepte egalas al $p(z_k)$. Por la aliaj z_k evidente ankaŭ la maldekstra flanko de (2.53.4) egalas al la informeca ekspluato, do al la dekstra flanko.

Se oni do adicias ambaŭflanke por ĉiuj k, tiam pluvalidas la <-signo, kaj rezultas

$$\sum_k \sum_i p(y_i)\cdot p(z_k \mid y_i)\cdot \mathrm{ld}\frac{1}{p(z_k \mid y_i)} < \sum_k p(z_k)\cdot \mathrm{ld}\frac{1}{p(z_k)} \tag{2.53.7}$$

Die rechte Seite ist nach Definition gleich H(Z), die linke ist gleich

La dekstra flanko estas laŭdifine H(Z), la maldestra egalas al

$$\sum_i p(y_i)\cdot \sum_k p(z_k \mid y_i)\cdot \mathrm{ld}\frac{1}{p(z_k \mid y_i)} = \sum_i p(y_i)\cdot \mathrm{H}(Z \mid y_i)$$

also nach Definitionsgleichung (2.3.4) gleich H($Z|Y$). Die bewiesene Ungleichung (2.53.7) ist also identisch mit der zu beweisenden Aussage (2.53.1). -

Im Falle stochastischer Unabhängigkeit zwischen Y und Z gibt es *kein* z_k, für welches die Ungleichungen (2.53.4) und (2.53.6) gelten. Vielmehr stimmen dort für *alle* z_k die beiden Seiten überein, folglich auch in der Ungleichung (2.53.7). Für deren rechte Seite gilt (wegen der Übereinstimmung aller Felder $Z|y_k$, also aller Unsicherheiten) H($Z|Y$) = H(Z). Damit geht (2.53.7) in die Gleichung (2.53.2) über. -

Ebenfalls mittels der Jensenschen Ungleichung ist ferner beweisbar, dass

do laŭ la difinegalaĵo (2.3.4) egala al H($Z|Y$). La pruvita malegalaĵo (2.53.7) estas do identa kun la pruvenda aserto (2.53.1). -

Kaze de stokasta sendependeco inter Y kaj Z ne ekzistas z_k, por kiu validus la malegalaĵoj (2.53.4) kaj (2.53.6). Male ambaŭ flankoj egalas por *ĉiuj* z_k tie, sekve ankaŭ en la malegalaĵo (2.53.7). Por ĉi ties dekstra flanko validas (pro la identeco de ĉiuj kampoj $Z|y_k$, do de ĉiuj necertecoj) H($Z|Y$) = H(Z). Tiel transformiĝas (2.53.7) en la egalaĵon (2.53.2). -

Ankaŭ pere de la malegalaĵo de Jensen estas plue pruveble, ke

$$\mathrm{H}_Y(Z) := \mathrm{H}(Z|Y) \geq \mathrm{H}(Z|XY) =: \mathrm{H}_{XY}(Z) \tag{2.53.8}$$

Von hier aus beweist man (durch Einsetzen des Produktfelds XY für Y und des Vorgängerfelds W für X - und weiter durch vollständige Induktion) die schon als (1.2.9a,b) aufgestellte Vermutung. Im Spezialfall

Deirante de ĉi tio oni pruvas (enmetante la prudutkampon XY anstataŭ Y kaj la antaŭkampon W anstataŭ X – kaj plue per matematika indukto) la jam supozitan rilaton (1.2.9a,b). En la speciala kazo

$$Z=Z(t);\ Y=Z(t-1);\ X=Z(t-2);\ \ldots$$

bedeutet dies:

In einer Folge von Zeichen nimmt der Erwartungswert der Information des näch-

tio signifas:

En signosinsekvo la ekspekto de la informacio de la sekvanta signo des pli *sten Zeichens immer weiter ab (oder – im Falle stochastischer Unabhängigkeit – er bleibt gleich), je mehr Vorgängerzeichen schon bekannt sind. Dementsprechend fällt die Knappheit und wächst die Redundanz*

malpligrandiĝas (aŭ – kaze de stokasta sendependeco – restas konstanta), ju pli da antaŭirintaj signoj jam *estas konataj.* Konforme ŝrumpas la koncizeco kaj kreskas la redundo

$$r_n := 1 - k_n := 1 - H_n/\mathrm{ld}\ u \qquad (2.53.9)$$

wobei n die Zahl der Vorgängerzeichen ist.

kie n signas la nombron de antaŭirintaj signoj.

Daraus darf nicht vorschnell geschlossen werden, der Grenzwert H_∞ (also auch die Knappheit k_∞) sei 0, die Redundanz r_∞ betrage 100%, das Zeichen sei also durch seine Vorgeschichte festgelegt. Wenn die Zeichenquelle *nicht* (wie ein Rechner, der zur Erzeugung einer – scheinbaren! - Zufallszeichenfolge programmiert wurde) ein *deterministisches* System ist, sondern ein problistisches (falls es ein solches in der Realität überhaupt gibt, was der philosophische Determinismus bestreitet), gilt der Satz mit $H_\infty > 0$, $k_\infty > 0$, $r_\infty < 1$.

El tio oni ne tro haste konkludu, la limeso H_∞ (do ankaŭ la koncizeco k_∞) estus 0, la redundo r_∞ 100%a, la signo do determinita fare de sia antaŭhistorio. Se la fonto de la signoj *ne* estas (kiel komputilo programita por produkti – fikcian! - vicon de hazardaj signoj) *determinisma* sistemo, sed sistemo probabilisma (se tia entute ekzistas en la realeco, kion kontestas la filozofia determinismo), tiam la teoremo validas kun $H_\infty > 0$. $k_\infty > 0$, $r_\infty < 1$.

2.54 *Subjektive Information und subjektive Unsicherheit.*

2.54 *Subjektiva informacio kaj subjektiva necerteco.*

Mit Gleichung (2.11.1) hatten wir die Information von vornherein auf einen Empfänger in einer bestimmten Situation bezogen und dies durch die Abhängigkeit von der „subjektiven Wahrscheinlichkeit“ w zum Ausdruck gebracht. Zwei verschiedene Empfänger (Beobachter) A und B können dieselbe Nachricht (dasselbe Zeichen oder sonstige Ereignis) mit verschiedener Wahrscheinlichkeit erwarten. A möge die Quelle schon so lange beobachtet haben, dass er durch Wahrscheinlichkeitslernen – z.B. nach Gleichung (2.18.1) - gut an ihre statistischen Eigenschaften angepasst ist. *Seine* subjektiven Wahrscheinlichkeiten p_i stimmen daher mit den Auftrittshäufigkeiten h_i der Zeichen (Ereignisse) gut überein. Die Unsicherheit H des Empfängers A wird daher auch *„objektive* Unsicherheit“ oder

Per la egalaĵo (2.11.1) ni dekomence rilatigis la informacion al ricevonto en certa situacio, kaj ni esprimis tion per la dependigo de la „subjektiva probablo“ w. Du diversaj ricevontoj (observantoj) A kaj B povas atendi la saman mesaĝon (la saman signon aŭ alian eventon) laŭ diversaj probabloj. A estu jam observinta la fonton tiom longe, ke li estas per probablolernado – ekzemple laŭ la egalaĵo (2.18.1) – bone adaptiĝinta al ĝia statistikaj trajtoj. *Liaj* subjektivaj probabloj p_i tial bone konformas kun la aperoftecoj h_i de la signoj (eventoj). La necerteco H de la ricevonto A tial estas ankaŭ nomita „*objektiva* necerteco“ aŭ „aritma informacio *de la fonto*“; konforme al tio oni nomas la pro-

„mittlere Information *der Quelle*" genannt; entsprechend bezeichnet man die Wahrscheinlichkeiten in seinem subjektiven Feld als „*objektive* Wahrscheinlichkeiten p_i." B dagegen apperzipiere aufgrund eines Feldes mit zwar gleichem Alphabet aber anderer Wahrscheinlichkeitsverteilung $w = (w_1, w_2, ..,. w_u) \neq p$, so dass für mindestens einen Index i $w_i \neq p_i$ ist. Die Information

bablojn en lia subjektiva kampo „*objektivaj* probabloj p_i". Male B aperceptu surbaze de kampo havanta ja la saman alfabeton sed alian probablodstribuon $w = (w_1, w_2, ..., w_u) \neq p$,, tiel ke por almenaŭ unu indico i validas $w_i \neq p_i$. Plue nomitas la informacio

(2.54.1) $$i_{\text{sub}} := \text{ld}\frac{1}{w_i}\text{bit}$$

eines Zeichens (Ereignisses) für B heißt weiterhin „*subjektive* Information". Diese subjektive Information einer Nachricht für B ist größer als die Information für den an die Quelle angepassten Empfänger A, falls $w_i < p_i$, wenn also der Empfänger A diese Nachricht eher vorausgesehen hatte als B. Wegen der allgemeingültigen Gleichung (2.12.5) muss es dann aber zwangsläufig auch ein $w_i > p_i$ geben, also eine Nachricht, die fur B wahrscheinlicher ist als für A, so dass in den - gemessen an der höheren Erwartung von B - seltenen Fällen, in denen diese Nachricht eintrifft, B eine geringere „subjektive Information" erhält als A. *Im Mittel* lohnt es sich aber, sich an die Wahrscheinlichkeitsverteilung der Quelle anzupassen. Denn der angepasste Empfänger A hat *im Mittel* weniger Information zu apperzipieren als der nicht angepasste Empfänger B. Genauer gilt folgender, mittels der Jensenschen Ungleichung zu beweisende

de signo (evento) por B „*subjektiva* informacio". Ĉi tiu subjektiva informacio de mesaĝo por B estas pli granda ol la informacio por la al la fonto pli bone adaptita ricevonto A, se $w_i < p_i$, se do la ricevonto A estas antaŭvidinta ĉi tiun mesaĝon pli verŝajne ol B. Pro la ĝenerale valida egalaĵo (2.12.5) nepre ekzistas iu $w_i > p_i$, do mesaĝo, kiu estas por B pli probabla ol por A, tiel ke en la – laŭ la pli alta ekspekto de B – maloftaj kazoj, en kiuj ĉi tiu mesaĝo alvenas, B ricevas malpli grandan „subjektivan informacion" ol A. *Mezume* tamen indas, adaptiĝi al la probablodistribuo de la fonto. Ĉar la adaptiĝinta ricevonto A devas *mezume* apercepti malpli da informacio ol la ne adaptiĝinta ricevonto B. Kiel pruveblas per la JENSENA malegalaĵo, validas la pli ekzakta

Satz: Falls zwei Felder im Alphabet übereinstimmen, jedoch wenigstens ein Element im ersten Feld die Wahrscheinlichkeit p_i, im zweiten die Wahrscheinlichkeit $w_i \neq p_i$ hat, gilt

Teoremo: Se du kampoj koincidas en la alfabeto, sed se almenaŭ unu elemento en la unua kampo havas la probablon p_k, en la dua la probablon $w_i \neq p_i$,, tiam validas

(2.54.2) $$H_{\text{sub}}/\text{bit} := \sum_i p_i \text{ld}\frac{1}{w_i} > \sum_i p_i \text{ld}\frac{1}{p_i} = H/\text{bit}$$

Beweis: ld x ist eine in]0, ∞[konvexe Funktion, erfüllt also dort die JENSEN-

Pruvo: ld x estas funkcio konveksa en]0, ∞[, do tie plenumas la JENbedingung (a). Ferner erfüllen die Quotienten w_i/p_i die Argumentbedingungen (b, c), denn es handelt sich um positive Zahlen[28], von denen nach den soeben angestellten Überlegungen mindestens eine größer als 1, und mindestens eine kleiner als 1 ist. Als Gewichtszahlen g_i benutzen wir die p_i. Die Jensensche Ungleichung (2.50.4) nimmt damit die spezielle Form

SENan kondiĉon (a). Krome la kvocientoj w_i/p_i plenumas la kondiĉojn (b, c), ĉar temas pri pozitivaj nombroj[28], de kiuj laŭ la ĵusaj konsideroj almenaŭ unu superas 1, kaj almenau unu ĝin malsuperas. Kiel pezoj g_i ni uzas la p_i. Tiel la JENSENA malegalaĵo (2.50.4) transformiĝas speciale en

$$\sum_i p_i \cdot \mathrm{ld}\frac{w_i}{p_i} < \mathrm{ld}\left(\sum_i p_i \cdot \frac{w_i}{p_i}\right) = \mathrm{ld}\left(\sum_i w_i\right) = \mathrm{ld}\,1 = 0 \tag{2.54.3}$$

an. Die linke Seite der Ungleichung ist also negativ. Der dortige Ausdruck lässt sich aufspalten in

La maldekstra flanko de la malegalaĵo do estas negativa. Oni povas disfendi la tiean esprimon en

$$\sum_i p_i \cdot \mathrm{ld}\frac{1}{p_i} + \sum_i p_i \cdot \mathrm{ld} w_i = \sum_i p_i \cdot \mathrm{ld}\frac{1}{p_i} - \sum_i p_i \cdot \mathrm{ld}\frac{1}{w_i}$$

Da der Ausdruck nach (2.54.3) negativ sein muss, muss H_{sub} größer als H sein. Dies war zu beweisen.

Ĉar la esprimo devas esti laŭ (2.54.3) negativa, H_{sub} devas superi H. Tio estis pruvenda.

In der allgemeinen Ungleichung

En la ĝenerala malegalaĵo

$$H^{\mathrm{sub}}/\mathrm{bit} = \sum_i p_i \ \mathrm{ld}\frac{1}{w_i} \geq \sum_i p_i \ \mathrm{ld}\frac{1}{p_i} = H/\mathrm{bit} \tag{2.54.4}$$

gilt wegen des bewiesenen Satzes das Gleichheitszeichen dann und nur dann, wenn alle $w_i=p_i$ sind, also die Felder für A und B voll übereinstimmen. -

validas pro la pruvita teoremo la egalecsigno tiam kaj nur tiam, kiam ĉiuj $w_i=p_i$, se do la kampoj plene koincidas por A kaj B. -

Zweierlei ist bei der Interpretation der Ungleichungen (2.54.2) und (2.54.4) zu beachten.

Oni atentu, interpretante la malegalaĵojn (2.54.2) kaj (2.54.4), du faktojn.

Erstens muss H_{sub} nicht gleich der Unsicherheit des Empfängers B sein. Dessen (subjektive) Unsicherheit H^{sub} berechnet sich (vgl. Kapitel 1.3 und 2.3) mit (2.19.3) aus der Wahrscheinlichkeitsverteilung w des *seine* Situation beschreibenden Feldes zu

Unu H_{sub} ne nepre egalas al la necerteco de la ricevonto B. Ties (subjektiva) necerteco H^{sub} kalkuliĝas (laŭ la ĉapitroj 1.3 kaj 2.3) per (2.19.3) el la probablo-distribuo w de la kampo, kiu priskribas *lian* situacion, do kiel

[28] Wir setzen voraus, dass alle p_i positiv sind, dass also B nicht abergläubisch ist (etwas Unmögliches für möglich hält), und dass ferner alle w_i positiv sind, also B nichts Mögliches für ein Wunder hält. Die Ungleichung (2.54.2) gilt aber auch für diese beiden Sonderfälle. Ein einfacher Beweis dafür findet sich in (Frank 1964b).

[28] Ni supozu, ke ĉiuj p_i estas pozitivaj, do ke B ne estas superstica (kredante ion ne eblan ebla), kaj ke plue ĉiuj w_i estas pozi-tivaj, do B ne vidas en io ebla miraklon. Sed la malegalaĵo (2.54.2) validas ankaŭ en tiuj apartaj kazoj. Simpla pruvo por tio troviĝas en (Frank, 19664b).

(2.54.5) $$H^{\text{sub}} / \text{bit} := \sum_i w_i \cdot \text{ld} \frac{1}{w_i}$$

Sie kann größer, gleich oder kleiner sein als die Unsicherheit *H* von A. (Wer eine Nachrichtenquelle - beispielsweise das Verhalten eines gefährlichen Raubtieres - nicht kennt, dessen Unsicherheitsgefühl braucht nicht besonders gross zu sein - er kann auch ein trügerisches Gefühl besonders *geringer* Unsicherheit haben! Ist umgekehrt für B jede mögliche Nachricht der Quelle gleichwahrscheinlich, ist seine Unsicherheit maximal, also größer als die von A. Schließlich ist die Unsicherheit von A und B dieselbe, wenn die Wahrscheinlichkeitsverteilung *w* nur eine Vertauschung der Wahrscheinlichkeiten von *p* ist.) H_{sub} ist vielmehr[29] der (aus der Sicht von A geltende) Erwartungswert - der theoretische Mittelwert - der subjektiven Information i_{sub}, die B mit einer Nachricht von der Quelle erhält. Er kann größer, kleiner oder gleich der subjektiven Unsicherheit von B sein – aber, wie bewiesen, nicht kleiner als die Unsicherheit von A.

Zweitens wurde im Beweis die Voraussetzung gar nicht benutzt, A sei an die Quelle angepasst (zumindest besser als B), seine Unsicherheit und seine subjektiven Informationswerte seien also (wenigstens in besserer Näherung) die „objektiv“ bestehenden. Der bewiesene Satz besagt also allgemein: Für jeden Nachrichtenempfänger ist *sein Erwartungswert* der subjektiven Information einer Nachricht für einen *anderen Empfänger größer* als die *eigenen* Unsicherheit (oder zumindest gleich groß wie diese).

Übungsaufgabe 2.5(1)

Wie groß ist die absolute und wie groß die realtive Redundanz, und wie groß ist die

[29] Zur Vermeidung des Missverständnisses sollte für H_{sub} das Wort „Unsicherheit“ vermieden werden.

Ĝi povas esti pli granda, egala aŭ malpli granda ol la necerteco *H* de A. (Kiu ne konas informfonton – ekzemple la konduton de danĝera rabbesto - ties sento de necerteco ne nepre estas aparte forta – li povas havi ankaŭ la trompan senton de aparte *malgranda* necerteco! Se inverse por B ĉiu ebla mesaĝo de la fonto samprobablas, tiam lia necerteco estas maksimuma, do pli granda ol tiu de A. Fine la necerteco de A kaj B egalas, se la probablodistribuo *w* estas nur permutaĵo de la probabloj de *p*.) H_{sub} estas pli ĝuste[29] la ekspekto – la teoria aritmo – (el la perspektivo de A) de la subjektiva informacio i_{sub}, kiun ricevas B per mesaĝo de la fonto. Ĝi povas esti pli granda, malpli granda aŭ egala al la subjektiva necerteco de B – sed, kiel pruvite, ne malpli granda ol la necerteco de A.

Due en la pruvo tute ne estas uzita la premiso, ke A estus adaptiĝinta (almenaŭ pli bone ol B) al la fonto, do liaj subjektivaj informacivaloroj estus (almenaŭ kiel pli bonaj proksimumaĵoj) la „objektive“ validaj. La pruvita teoremo do asertas ĝenerale: Por ĉiu ricevonto de mesaĝo estas *lia ekspektvaloro* de la subjektiva informacio, kiun havas mesaĝo *por alia ricevonto, pli granda* ol la *propra* necerteco (aŭ almenaŭ egala al ĉi tiu).

Ekzerctasko 2.5(1)

Kiom granda estas la absoluta kaj kiom la relativa redundo, kaj kiom

[29] Cele eviton de miskomprenoj oni evitu por H_{sub} la vorton „necerteco“.

Knappheit der Vokalfolge, die im propädeutischen Kapitel 1.2 (Seite 16) als Beispiel benutzt wurde, wenn man
1. die (sehr wahrscheinliche!) stochastische Abhängigkeit in der Folge nicht berücksichtigt?
2. die stochastische Abhängigkeit berücksichtigt und voraussetzt, dass die bedingten Wahrscheinlichkeiten mit den bedingten relativen Häufigkeiten identisch sind?

grandas la koncizeco de la vokalsinsekvo, uzita en la propedeŭtika ĉapitro 1.2 (paĝo 16) kiel ekzemplo, se oni
1. ne konsideras la (tre probablan!) stokastan dependecon en la sinsekvo?
2. ja konsideras la stokastan dependecon kaj supozas la kondiĉitajn probablojn identaj kun la kondiĉitaj relativaj oftecoj?

Übungsaufgabe 2.5(2)

Ein Leser der $N = 32$ Zeichen langen Vokalfolge aus Kapitel 1.2 apperzipiert sie aufgrund der subjektiven Wahrscheinlichkeiten w_k, die unter sich alle übereinstimmen. Wie groß ist seine subjektive Unsicherheit H^{sub}? Wie groß ist der Erwahrtungswert H_{sub} der subjektiven Information, die ihm das nächste Vokalzeichen bringt?

Ekzerctasko 2.5(2)

Leganto de la $N = 32$ signojn longa vokalsinsekvo el ĉapitro 1.2 aperceptas ĝin surbaze de subjektivaj probabloj w_k inter si ciuj egalaj. Kiom estas lia subjektiva necerteco H^{sub} ? Kiom estas la ekspektvaloro H_{sub} de la subjektiva informacio, kiun havigos al li la sekvanta vokalsigno?