Aufgabe: Wie ähnlich sind sich zwei Menschen mindestens, wenn wir annehmen, daß 75 % aller autosomalen Gene nur ein einziges Allel besitzen und die relative Häufigkeit polymorpher Gene einer unendlichen geometrischen Reihe folgt?

Lösung: Sei h_i die relative Häufigkeit, ein Gen mit *i* Allelen anzutreffen, und $h_1 = 0{,}75$. Sei ferner

$$h_i = h_1 q^{i-1} \quad \text{für} \quad i \in \{1, \ldots, n\}$$

die entsprechende geometrische Folge und

$$\sum_{i=1}^{n} h_i = \sum_{i=1}^{n} h_1 q^{i-1} = h_1\left(1 + q + \ldots + q^{n-1}\right) = h_1 \frac{1-q^n}{1-q}$$

die zugehörige endliche geometrische Reihe. Für $n \to \infty$ geht diese Reihe in die unendliche geometrische Reihe

$$\sum_{i=1}^{\infty} h_i = h_1 \frac{1}{1-q}$$

über. Diese Erweiterung können wir ohne Beschränkung der Allgemeinheit vornehmen, da der Beitrag höherer Glieder verschwindend gering ist. Aus der Normierungsbedingung

$$\sum_{i=1}^{\infty} h_i = 1$$

folgt daraus $q = 0{,}25.$ Damit ergibt sich die geometrische Folge zu

$$h_i = h_1 \frac{1}{4^{i-1}} = \frac{3}{4^i} \quad \text{für} \quad i \in \{1, \ldots, n\}.$$

Da der Mensch ungefähr 22500 Gene hat, ist für

$$i = \frac{\ln 3 + \ln 22500}{\ln 4} \approx 8$$

die relative Häufigkeit

$$h_8 = \frac{3}{4^8} = 4{,}58 \cdot 10^{-5} \approx \frac{1}{22500}$$

erreicht, bei der maximal noch ein Gen einen Beitrag leistet. Vergleicht man nun zwei Individuen, so ist immer eine die Referenzperson.

Stimmen beide Personen in allen 22500 Genen überein wie etwa eineiige Zwillinge, so beträgt ihre Ähnlichkeit 100 %. Ist nur ein Gen bei beiden unterschiedlich, so liegt ihre Ähnlichkeit immer noch fast bei 1:

$$\frac{N-\Delta N}{N}=1-\frac{\Delta N}{N}=1-\frac{22500-16875}{5625\cdot 22500}=1-\frac{1}{22500}=99{,}996\ \%.$$

Nehmen wir an, daß sich zwei Personen statistisch in jedem zweiten Gen unterscheiden, so sind von den 5625 polymorphen Genen ungefähr 2812 unterschiedlich und die Ähnlichkeit liegt bei

$$1-\frac{1}{2}\frac{5625}{22500}=87{,}5\ \%.$$

Die größte Unähnlichkeit bzw. geringste Ähnlichkeit liegt vor, wenn sie sich in jedem Gen unterscheiden würden. Diese betrüge dann

$$1-\frac{5625}{22500}=75\ \%,$$

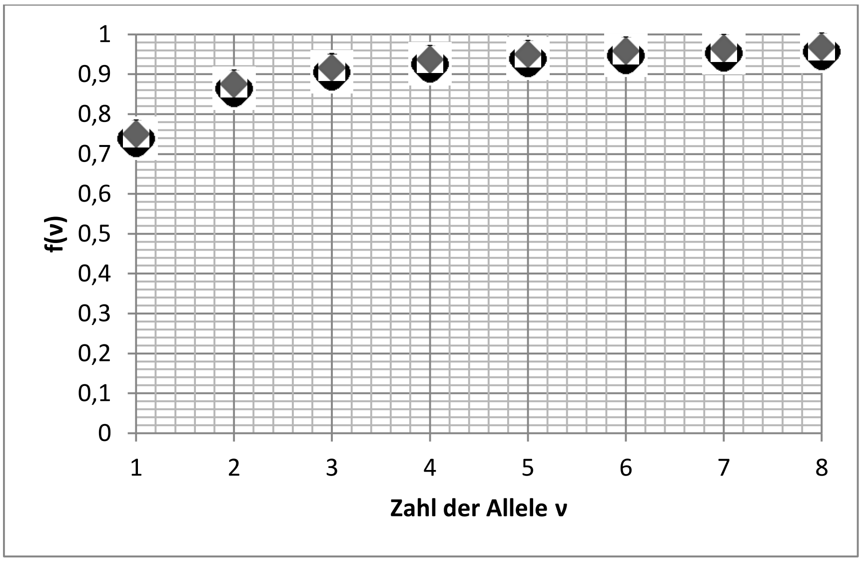
entspräche also exakt dem prozentualen Anteil monomorpher Gene. Unähnlicher können Menschen unter diesen Annahmen nicht sein. Engere Verwandte, die sich beispielsweise nur in jedem dritten polymorphen Gen unterscheiden, hätten demnach eine Ähnlichkeit von

$$1-\frac{1}{3}\frac{5625}{22500}=91{,}67\ \%$$

bzw. allgemein

$$f(\nu)=1-\frac{1}{\nu}\frac{5625}{22500},$$

wobei $\nu\in\{1,\ldots,5625\}$. Diese Funktion ist nachfolgend graphisch dargestellt:



In Wirklichkeit kann aber die Ähnlichkeit einen etwas größeren Minimalwert nicht unterschreiten, denn die Zahl der Homozygoten in einem Gen mit *i* Allelen ist gleich *i*, während die Zahl der Genotypen gleich i^2 ist. Im Verhältnis ist also die gewichtete Ähnlichkeit um 1/*i* kleiner als h_i. Sei also *N* die oben angegebene Zahl der menschlichen Gene und

$$N_i = h_i N$$

die absolute Häufigkeit, ein Gen mit *i* Allelen anzutreffen. Wir berechnen daraus nun die Wahrscheinlichkeit w_i, daß zwei Vergleichspersonen das gleiche Allel besitzen. Diese errechnet sich als die gewichtete relative Häufigkeit, bei *i* homozygoten Phänotypen ein ähnliches Gen anzutreffen,

$$w_i = \frac{h_i}{i} = \frac{1}{i}\frac{N_i}{N}.$$

In der nachfolgenden Tabelle haben wir alle Ergebnisse noch einmal zusammengefaßt:

i	h_i	N_i	w_i
1	0,75	16875	0,75
2	0,1875	4219	0,093755556
3	0,046875	1055	0,01562963
4	0,01171875	264	0,002933333
5	0,00292969	66	0,000586667
6	0,0007342	16	0,000118519
7	0,00018311	4	$2{,}53968 \cdot 10^{-5}$
8	$5{,}5776 \cdot 10^{-5}$	1	$5{,}555568 \cdot 10^{-6}$
Σ	0,99998474	22500	0,863054656

Danach ist die gewichtete Häufigkeit, ein ähnliches Gen mit *i* Allelen anzutreffen, stets kleiner oder gleich der Häufigkeit, das Gen selbst anzutreffen. Die genetische Ähnlichkeit einschließlich aller homozygoten Allelkombinationen muß daher größer als 75 % sein, beträgt aber mindestens

$$\sum_{i=1}^{8} w_i = 86{,}31\,\%.$$

Das ist unter den getroffenen Annahmen die Wahrscheinlichkeit, mit der sich auch zwei sonst grundverschiedene Menschen ähnlich sind. Sie kann nur überschritten, niemals aber unterschritten werden.

