$n!$ közelítése

Eszköz

Euler

$$\sum_{h=m}^{n-1} f(h) = \\ = \int_m^n f(t)\,\mathrm{d}t - \tfrac{1}{2}\big(f(n)-f(m)\big) + \sum_{k=1}^{K-1} \frac{B_{2k}}{(2k)!}\big(f^{(2k-1)}(n)-f^{(2k-1)}(m)\big) + R_K$$

B_k a Bernoulli-számok: $B_0 = 1$, $m > 0$-ra $B_k = -\dfrac{1}{k+1}\displaystyle\sum_{h=0}^{k-1}\binom{k+1}{h}B_h$;

$$R_K = R_{f,m,n,K} = \frac{B_{2K}}{(2K)!}\sum_{h=m}^{n-1} f^{(2K)}(h+\vartheta_h), \qquad \vartheta_h = \vartheta_{f,K,h} \in [0,1].$$

Ha az $[m,n]$ intervallumban $f^{(2k)}$ és $f^{(2k+2)}$ előjele minden k-ra megegyezik:

$$R_K = \vartheta_K \frac{B_{2K}}{(2K)!}\big(f^{(2K-1)}(n) - f^{(2K-1)}(m)\big), \qquad \vartheta_K = \vartheta_{f,m,n,K} \in [0,1].$$

Stirling

$$\log n! = \log\Gamma(n+1) = \\ = n\log n - n + \tfrac{1}{2}\log n + \tfrac{1}{2}\log(2\pi) + \sum_{k=1}^{K-1}\frac{B_{2k}}{2k(2k-1)}n^{-(2k-1)} + R_K$$

$$R_K = \vartheta_K\frac{B_{2K}}{2K(2K-1)}n^{-(2K-1)}, \qquad \vartheta_K \in [0,1];$$

$$R_1 \in \left[0, \frac{1}{12n}\right], \quad R_2 \in \left[\frac{-1}{360n^3}, 0\right], \quad R_3 \in \left[0, \frac{1}{1260n^5}\right], \quad R_4 \in \left[\frac{-1}{1680n^7}, 0\right].$$

Valós $n>0$-ra, sőt komplex $\mathrm{Re}\,n>0$-ra is jó; nagyobb $\mathrm{Re}\,n$-re pontosabb. Minden n-re igaz (a pólusokban is), hogy $(n-1)! = \frac{n!}{n}$ és $(-n)! = \dfrac{\pi}{\sin\pi n}\dfrac{n}{n!}$.

[Részletes levezetés található például A. O. Gelfandnál: *Исчисление конечных разностей*; magyarul is van: *Differenciaszámítás*, Akadémiai Kiadó, 1954, de nagyon sajtóhibás.]

Banach gyufásdobozai

Rényi: — Banach professzor szenvedélyes dohányos volt; hogy ne kelljen soká keresgélnie gyufa után, mindig két doboz gyufát tartott magánál: egyet a jobb zsebében, egyet pedig a bal zsebében. Amikor rá akart gyújtani, hol a bal, hol pedig a jobb zsebéből vette elő a gyufát, fele-fele valószínűséggel. Egyik nap két tele doboz gyufát tett a zsebébe, egy tele dobozban n szál gyufa van. Amikor valamelyik dobozt üresen találta, várhatólag hány szál gyufa volt még a másik dobozban?

Jelöljük $P_k(n)$-nel annak a valószínűségét, hogy addigra a másik dobozból k szál fogyott (vagyis $n-k$ maradt benne). Könnyen ellenőrizhető, hogy

$$P_k(n) = \binom{n+k}{k} 2^{-(n+k)}.$$

Így a várható érték

$$M(n) = \sum_{k=0}^{n} (n-k) P_k(n).$$

Ez rövidebb alakra hozható:

$$M(n) = (2n+1)\binom{2n}{n} 2^{-2n} - 1.$$

A legegyszerűbben talán így lehet erről megbizonyosodni: Adjunk hozzá mindkét képlethez 1-et, mégpedig úgy, hogy a szummához tagonként hozzáadjuk $\sum_{k=0}^{n} P_k(n)$-et (ami 1, lévén egy teljes eseményrendszer valószínűségeinek összege), azután lássuk be, hogy $n=1$-re (vagy akár $n=0$-ra) fennáll az egyenlőség, a továbbiakban pedig $(2n+2)\big(M(n+1)+1\big) = (2n+3)\big(M(n)+1\big)$ mind a szumma, mind a rövid alak esetében.

Mennyi az ennyi? A Stirling-sort a B_4-es tagig használva

$$\binom{2n}{n} 2^{-2n} = \frac{1}{\sqrt{\pi n}} \exp\big(-\tfrac{1}{8}n^{-1} + \tau_n n^{-3}\big),$$

ahol $\tau_n \in \big[-\frac{1}{2880}, \frac{1}{180}\big]$, $|\tau_n| \le \frac{1}{180}$.

Most szorozzunk be $2n+1$-gyel, emeljük ki szorzónak a főtagot, fejtsük sorba az exp-et, csonkítsunk, és végül vonjunk le 1-et:

$$M(n) = \frac{2}{\sqrt{\pi}} \sqrt{n}\big(1 + \tfrac{3}{8}n^{-1} - \tfrac{7}{128}n^{-2} + \sigma_n n^{-3}\big) - 1;$$

itt a hibatagok végigkövetéséből látható, hogy $|\sigma_n| < 0.01$.

$\frac{2}{\sqrt{\pi}} = 1.128\,379\,167\,095\,512\,5746^{-}$.

Jelöljük $\tilde{M}(n)$-nel a hibatag elhagyásával adódó képletet. Explicit kiszámolás azt mutatja, hogy $\tilde{M}(n)$ hibája már $n=1$-re is kisebb 0.0102-nél;

$\tilde{M}(40) = 6.203\,157\,196^{-}$ vs. $M(40) = 6.203\,158\,181^{-}$,
$\tilde{M}(100) = 10.326\,044\,181^{+}$ vs. $M(100) = 10.326\,044\,281^{-}$.

[*Variáns*: Hány szál várható a másik dobozban, amikor az egyik doboz *kiürül*? Most $P_k(n) = \binom{n-1+k}{k} 2^{-(n-1+k)}$, $M(n) = (2n-1)\binom{2n-2}{n-1} 2^{-(2n-2)}$; $M(n+1) - 1 =$ az előbbi $M(n)$. Nem folytatom.]

Kettős születésnapok

Az eredeti kérdésfeltevés így szól: hány találomra kiválasztott ember kell ahhoz, hogy legalább fele valószínűséggel legyen közöttük kettő, akinek ugyanarra a napra esik a születésnapja. (A válasz: 23.)

Valamivel könnyebben kezelhető a kérdés, ha fordítva fogalmazzuk meg: legfeljebb $\frac{1}{2}$ valószínűséggel ne legyen közöttük kettő.

Kicsit általánosítva a feladatot: tekintsük az n-féle jelből (pl. számjegyből, betűből, ismérvből) álló m hosszúságú sorozatokat — ezeknek mekkora hányada áll csupa különböző jelből, és ez a hányad milyen $m = m(n)$-ekre lesz legfeljebb $\frac{1}{2}$.

Az ilyen sorozatok száma n^m, a különböző jelekből állóké $n(n-1)\cdots(n-m+1)$; $g(n,m)$-mel jelölve az arányukat

$$g(n,m) = \prod_{h=0}^{m-1} \left(1 - \frac{h}{n}\right).$$

Ebből adott n-hez ki lehet ugyan számolni a keresett m-et, de tudunk jobbat is. Írjuk át $g(n,m)$-et így:

$$g(n,m) = \frac{n!}{(n-m)!} n^{-m}.$$

Ha $x!$-on $\Gamma(x+1)$-et és n^{-m}-en $\exp(-m \log n)$-et értünk, akkor $g(n,m)$-nek pólusa van a negatív egész n helyeken és ezenkívül minden valós (sőt minden komplex) n, m helyen értelmezve van. Feltehető tehát a kérdés, hogy milyen m-re lesz $g(n,m)$ *pontosan* $\frac{1}{2}$.

Jelöljük ezt a függvényt $M(n)$-nel: $g\big(n, M(n)\big) = \frac{1}{2}$.

A függvényt ebben az alakban keressük:

$$M(n) = an^{\frac{1}{2}} + b + cn^{-\frac{1}{2}} + O(n^{-1}).$$

Hogy ez lehetséges, azt még nem tudjuk — ki fog derülni, hogy igen.

Legyen $f(n,m) = \log g(n,m)$; most ez kell: $f\big(n, M(n)\big) = -\log 2$.

Stirling miatt

$$f(n,m) = -m - n\log\left(1 - \frac{m}{n}\right) + m\log\left(1 - \frac{m}{n}\right) - \frac{1}{2}\log\left(1 - \frac{m}{n}\right) + R,$$

ahol $R \in \left[\frac{-1}{12}\frac{m}{n(n-m)}, 0\right]$.

A hibabecslések alakjának egyszerűsítése végett érdemes előre tudni, hogy a képleteket kizárólag $m < An^{\frac{1}{2}}$-re fogjuk használni később megadandó A-val. Így $|R| < \frac{A}{4}n^{-\frac{3}{2}}$, ha $n > (1.5A)^2$. (Ki fog derülni, hogy ez nem megszorítás.)

A $\log\left(1 - \frac{m}{n}\right)$-eket sorbafejtve és a sorokat megfelelő helyen csonkítva

$$f(n,m) = -\frac{1}{2}\frac{m^2}{n} - \frac{1}{6}\frac{m^3}{n^2} + \frac{1}{2}\frac{m}{n} - \frac{1}{12}\frac{m^4}{n^3} + \frac{1}{4}\frac{m^2}{n^2} + S,$$

ahol az együttes hiba S; $|S|$ (bőven) $< An^{-\frac{3}{2}} + O(n^{-2})$.

Helyettesítsük be m helyébe $an^{\frac{1}{2}} + b + cn^{-\frac{1}{2}}$-t és válasszuk meg a, b, c-t úgy, hogy a legmagasabb tag $-\log 2$-t adjon, a rákövetkező két tag pedig essék ki:

$$a = \sqrt{2\log 2} = 1.177\,410\,022\,515\,474\,691^{+},$$
$$b = \frac{1}{2} - \frac{1}{6}a^2 = \frac{3-2\log 2}{6} = 0.268\,950\,939\,813\,351\,564^{-},$$
$$c = \frac{1}{8}a^{-1} - \frac{1}{72}a^3 = \frac{9-4\log^2 2}{72\sqrt{2\log 2}} = 0.083\,495\,268\,439\,623\,963^{+}.$$

$n \geq 4$-re A választható $a + \frac{b}{2} + \frac{c}{4}$-nek, ami $< 1.\overline{3}$. Az együttható-adjusztáláskor bekerült új $n^{-\frac{3}{2}}$-es tag együtthatója absz. < 0.070, a továbbiaké < 0.059 és < 0.031. Így — A-t tekintetbe véve — a |hibatag| $< 1.35n^{-\frac{3}{2}}$, ha $n \geq 4$.

Legyen tehát a közelítésünk

$$\tilde{M}(n) = \sqrt{2\log 2}\,n^{\frac{1}{2}} + \frac{3-2\log 2}{6} + \frac{9-4\log^2 2}{72\sqrt{2\log 2}}n^{-\frac{1}{2}}.$$

($\frac{1}{2}$ helyett más <1 valószínűségre hasonlóan megy.)

A becslésnek persze egész számnak kell lennie; eszerint fogadni akkor érdemes, ha $m \geq \lceil \tilde{M}(n) \rceil$.

Explicit kiszámolás azt mutatja, hogy $\lceil \tilde{M}(n) \rceil$ $n = 1, 2, 3$-ra is jó.

Még egy hibalehetőség van: Akármilyen pontos közelítés mellett is elképzelhető, hogy egész szám esik a pontos érték és a közelítés közé, és a $\lceil\ \rceil$ után egységnyi hiba marad. Nem tudom, bekövetkezik-e valaha (azt hiszem, igen). Explicit kiszámolás azt mutatja, hogy $n \in [1, 2^{32}-1]$-re $\lceil \tilde{M}(n) \rceil$ mindenütt pontos — akár a pontos együtthatókkal pontosan, akár az együtthatók fenti numerikus közelítéseivel és 16+64-bites lebegőpontos aritmetikával számolva —, így a hibaarány $\leq 2^{-32}$.

Hogy 5×10^9 embernek legalább $1/2$ valószínűséggel más és más ujjlenyomata legyen, ahhoz az kell, hogy $\approx 1.80 \times 10^{19}$-féle ujjlenyomatot lehessen megkülönböztetni — ha ugyan egyformán valószínűek; különben több kell.

$\tilde{M}(365) = 22.767\,708^{+}$ vs. $M(365) = 22.767\,690^{+}$, $1 - g(365, 23) = 0.507\,297^{+}$;
$\tilde{M}(366) = 22.798\,495^{+}$ vs. $M(366) = 22.798\,478^{-}$, $1 - g(366, 23) = 0.506\,323^{+}$.[1]

Kedvcsináló

Többes egybeesések valószínűségére is adható pontos képlet. (Találd meg.) Pl. hármas születésnapra 365 vagy 366 napos év esetén a fogadási küszöb 88. Hármas egybeesésre $\tilde{M}(n)$ ilyen alakban adható meg: $an^{\frac{2}{3}} + bn^{\frac{1}{3}} + c + dn^{-\frac{1}{3}} + \ldots$; itt pl. $a = 1.608\,146\,354\,157\,472\,240^{+} \approx \sqrt[3]{6\log 2}$, stb.[2]

---

[1] M. S. Klamkin és D. J. Newman, Extensions of the Birthday Surprise; *Journ. Comb. Theory*, **3** (1967) 279–282, csak egy tagot ad: helyes kitevővel, de az együtthatóját $\sqrt{2}\,\Gamma(3/2)$-re ($= 1.253\,314\,137\,315\,500\,251^{+}$) becsüli. (Ez 365-re sem jó, $\lceil 23.9^{+} \rceil$ adódik.)

[2] Klamkin–Newman képlete hármas születésnap esetére szintén helyes kitevőjű tagot ad, de $\sqrt[3]{6}\,\Gamma(4/3)$ együtthatóval ($= 1.622\,651\,459\,449\,668\,642^{-}$); ez sem jó (365-re sem: $\lceil 82.9^{-} \rceil$ adódik.)