STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
AFDELING ZUIVERE WISKUNDE

ZW 1965-002
S 339

BALLOT PROBLEMEN

Voordracht in de serie "Actualiteiten"
gehouden op zaterdag 30 januari 1965

door

F. Göbel

Januari 1965

Printed at the Mathematical Centre at Amsterdam,49,2nd Boerhaavestraat.
The Netherlands.

The Mathematical Centre, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Scientific Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Veronderstel dat bij een verkiezing waaraan twee kandidaten A en B deelnemen, kandidaat A een totaal van a stemmen behaalt, en B,b stemmen. Als $b \leq a$, wat is dan de kans p dat A gedurende het tellen van de stemmen steeds voor staat? We nemen aan dat de a+b stemmen één voor één en in aselecte volgorde worden geteld. Onder deze aanname is het antwoord op bovenstaande vraag van een verrassend eenvoudige vorm:

$$p = \frac{a - b}{a + b} . \tag{1}$$

Vaak wordt dit resultaat aan BERTRAND [1], of aan ANDRÉ [2] toegeschreven, maar volgens BARTON en MALLOWS [3] is het probleem door WHITWORTH al in 1878 gesteld en opgelost, en in 1886 opgenomen in zijn boek "Choice and Chance" [4].

Een eenvoudig bewijs kan worden gegeven met behulp van het zgn. spiegelingsprincipe. Bij de volgende meetkundige voorstelling komt dit principe het best tot zijn recht.

2. Stel a_m, resp. b_m is het aantal stemmen op A resp. B nadat m stemmen zijn geteld. De situatie $a_m = x$, $b_m = y$ kan worden voorgesteld door het punt (x,y) van een tweedimensionaal rooster van punten met gehele, niet-negatieve coördinaten. Een roosterpunt waarbij een situatie behoort die bij de telling voorkomt noemen we een <u>vertex</u>. Opvolgende vertices kunnen worden verbonden door lijnstukken (<u>stappen</u>) zodat een <u>pad</u> van (0,0) naar (a,b) ontstaat.

In deze meetkundige terminologie kan de zojuist gestelde vraag als volgt worden geformuleerd: hoe groot is het aantal paden van (0,0) naar (a,b) waarvan de vertices, afgezien van (0,0), onder de lijn x = y liggen?

Deze formulering leent zich goed tot generalisaties van het probleem; we zullen echter eerst het spiegelingsprincipe op de gestelde vraag toepassen.

3. Het aantal paden van (0,0) naar (a,b) is $\binom{a+b}{a}$. De paden waarvan de eerste stap verticaal is, voldoen geen van alle aan de gestelde eis;

hun aantal is $\binom{a+b-1}{a}$. Van de paden waarvan de eerste stap horizontaal is, en waarvan de vertices op of boven de lijn x = y liggen, is het aantal even groot. Immers, door het gedeelte tussen (0,0) en het eerste punt op de lijn x = y te spiegelen t.o.v. deze lijn ontstaat een pad van de eerste categorie. De kans dat een aselect pad aan de gestelde eis voldoet is dus

$$\frac{\binom{a+b}{a} - 2\binom{a+b-1}{a}}{\binom{a+b}{a}} = \frac{a-b}{a+b} .$$

4. De in paragraaf 2 gestelde vraag wijzigen we nu als volgt: in hoeverre ligt een aselect pad boven de lijn x = my? In principe vragen we hierbij naar de volledige kansverdeling van een nader te definiëren grootheid [1] $\underline{h}$ die de "hoeveelheid pad" boven de lijn x = my meet; soms beschouwen we alleen de eerste term van de verdeling. (Bij het bepalen van eerste doorgangstijden kan dit al voldoende zijn.) Voor $\underline{h}$ komen o.a. de volgende grootheden in aanmerking:

$\underline{h}_1$, het aantal vertices (strikt) boven de lijn x = my;

$\underline{h}_2$, het aantal vertices (strikt) boven de lijn x = my, plus het aantal vertices op deze lijn waarvan de voorganger boven de lijn ligt;

$\underline{h}_3$, het aantal vertices op of boven de lijn x = my;

$\underline{h}_4$, de lengte van het pad boven de lijn x = my.

Bij $\underline{h}_2$ en $\underline{h}_3$ wordt de oorsprong niet meegeteld.

Van de gevonden resultaten die in dit schema passen noemen we de volgende.

4.1 CHUNG en FELLER [5] bepaalden voor m = 1 de verdeling van $\underline{h}_2$ (ofwel van $\underline{h}_4$, daar voor m = 1 zowel $\underline{h}_2$ als $\underline{h}_4$ gelijk is aan het aantal stappen boven de lijn x = my). Interessant is dat voor a = b deze verdeling homogeen is op 0, 2, ..., 2a. HODGES [6] heeft hiervan een combinatorisch bewijs gegeven dat veel eenvoudiger is dan het bewijs van CHUNG en FELLER. Het geval a = b is eigenlijk het lastigste gedeelte van het resultaat:

---

1) Stochastische variabelen worden onderstreept.

de verdeling voor $a \neq b$ is op eenvoudige wijze af te leiden uit de homogene verdeling voor $a = b$.

Ora ENGELBERG [7] bepaalde, voor $a = b$ en $m = 1$, de verdeling van $\underline{h}_3$ (wat in dit geval op hetzelfde neerkomt als de verdeling van $\underline{h}_1$).

Met behulp van HODGES' constructie kan men eenvoudig nagaan dat voor m geheel en $a = mb$ de verdeling van $\underline{h}_4$ homogeen is op $0, \frac{a+b}{a}, 2\frac{a+b}{a}, \ldots, a\frac{a+b}{a}$. (zie [8]).

4.2 TAKÁCS [9] bepaalde voor m geheel en $a \geq mb + 1$ de verdeling van $\underline{h}_3$. Voor $a = mb + 1$ is deze verdeling homogeen op $1, 2, \ldots, a+b$. Hoewel dit laatste een speciaal geval is van een der resultaten in het eerder verschenen artikel TAKÁCS [10], wordt het toch afzonderlijk bewezen.

BARBIER [11] vond al in 1887 dat voor m geheel en $a \geq mb$ geldt:

$$P[\underline{h}_3 = 0] = \frac{a - mb}{a + b} \tag{2}$$

$$P[\underline{h}_1 = 0] = \frac{a + 1 - mb}{a + 1} \tag{3}$$

Bewijzen hiervan werden gegeven door AEPPLI [12] en door DVORETZKY en MOTZKIN [13].

4.3 Van de schaarse resultaten voor niet-noodzakelijk gehele m noemen we allereerst de merkwaardige formule van GROSSMAN [14]: als $a = mb$, dan is

$$P[\underline{h}_1 = 0] = \binom{a+b}{a}^{-1} \sum_{\varphi(k)} \frac{F_1^{k_1} F_2^{k_2} \cdots}{k_1!\, k_2! \cdots} \tag{4}$$

waarin de som zich uitstrekt over alle partities $1^{k_1} 2^{k_2} \ldots$ van $k = (a,b)$, en waarin $F_j = \frac{1}{j\alpha + j\beta}\binom{j\alpha + j\beta}{j\alpha}$, $\alpha = a/k$, $\beta = b/k$. Het bewijs hiervan werd door BIZLEY [15] gegeven.

TAKÁCS bewees in [16] dat voor $a > mb$ geldt

$$P[\underline{h}_3 = 0] = \frac{a}{a + b} \sum_{j=0}^{b} C_j \frac{\binom{b}{j}}{\binom{a + b - 1}{j}} \tag{5}$$

waarin $C_0 = 1$, en de overige C_j bepaald zijn door de betrekking

$$\sum_{j=0}^{b} c_j \frac{\binom{b}{j}}{\binom{[bm] + b - 1}{j}} = 0 \quad (b = 1, 2, \ldots). \tag{6}$$

Ook in [13] wordt enige aandacht besteed aan niet-gehele waarden van m.

Tenslotte vond TAKÁCS [10] dat voor $(a,b) = 1$ en $a = mb$, de verdeling van $\underline{h}_3$ homogeen is op $1, 2, \ldots, a+b$. Hieruit volgt meteen (door "het plaatje" ondersteboven te houden) dat $\underline{h}_1$ homogeen is op $0, 1, \ldots, a+b-1$. Het is daarom verrassend om in bovengenoemd artikel te lezen: "It is very surprising that the probability that A has a majority of at least a : b throughout the counting is the same as the probability that his majority is less than a : b throughout the counting, except in the final record".

5. Natuurlijk kan het in paragraaf 2 geformuleerde probleem in vele andere richtingen worden gegeneraliseerd. Men kan het aantal kandidaten uitbreiden, men kan de lijn $x = my$ vervangen door een andere rechte of door een nog algemenere grenslijn (zie paragraaf 6), men kan vragen naar de kans dat een pad tussen twee gegeven rechten ligt (zie bijv. [10]). Een minder voor de hand liggende uitbreiding, eveneens afkomstig van TAKÁCS [16] is de volgende stelling.

Stelling: Een vaas bevat n kaarten die voorzien zijn van de niet-negatieve getallen $k_1, \ldots, k_n$. Stel $k_1 + \ldots + k_n = k$ en $0 \leq k \leq n$. De kaarten worden één voor één zonder teruglegging getrokken. Zij $\underline{\nu}_j$ het getal op de j^e kaart $(j = 1, \ldots, n)$. Dan geldt

$$P\left[\underline{\nu}_1 + \ldots + \underline{\nu}_r < r \quad \text{voor } r = 1, \ldots, n\right] = 1 - \frac{k}{n}. \tag{7}$$

Neemt men a kaarten met een 0, en b kaarten met het getal $(m+1)$, en wel zo dat $b(m+1) \leq a+b$, dan geeft toepassing van deze stelling direct formule (2).

Relaties als (7) kunnen, door hun algemene karakter, op allerlei gebieden worden toegepast, zoals wachttijden, stuwmeren, toetsingstheorie, Markov-ketens, "order statistics" en fluctuatie-theorie.

6. Stel $f_0, f_1, \ldots$ is een monotoon niet-dalende reeks van natuurlijke getallen, en stel N_m is het aantal paden van $(0,0)$ naar (m,f_m) met de eigenschap dat (m,f_m) het eerste punt van de gedaante (i,f_i) is dat op het pad ligt. Een dergelijk pad noemen we een <u>toegestaan</u> pad naar (m,f_m).

<u>Stelling</u>: N_m voldoet aan de recurrentiebetrekking

$$\sum_{i=0}^{m} (-1)^i N_i \binom{f_i}{m-i} = 0 \quad (m > 0). \tag{8}$$

<u>Bewijs</u>: Beschouw een willekeurig pad van $(0,0)$ naar (m,f_m). De rechter-eindpunten van de horizontale stappen vormen dan een monotoon niet-dalende rij van niet-negatieve getallen y_i $(i = 1, \ldots, m)$. Als het pad toegestaan is, geldt bovendien

$$y_i < f_{i-1} \quad (i = 1, \ldots, m). \tag{9}$$

Definiëer nu voor $0 \leq j \leq m$ een j-reeks als een reeks van niet-negatieve getallen $y_0 = 0, y_1, \ldots, y_m$ die voldoen aan (9), en bovendien aan

$$y_0 \leq y_1 \leq \ldots \leq y_j \tag{10}$$

$$y_j > y_{j+1} > \ldots > y_m. \tag{11}$$

Laat voor vaste m, A_j het aantal j-reeksen zijn. Nu kan voor $j = 0$, nooit voldaan zijn aan de eerste ongelijkheid van (11), zodat $A_0 = 0$. Het is ook direct te zien dat $A_m = N_m$ (mits we ook $y_0 = 0$ definiëren voor de reeksen die met een toegestaan pad corresponderen). Verder is het aantal j-reeksen plus het aantal (j+1)-reeksen, blijkens de gegeven definitie, gelijk aan het aantal reeksen van niet-negatieve getallen met $y_0 = 0$, die voldoen aan (9), (10) en

$$y_{j+1} > \ldots > y_m. \tag{12}$$

Dit laatste aantal kan men echter ook rechtstreeks bepalen, door op te merken dat voor de waarden $y_0, \ldots, y_j$ het aantal mogelijkheden N_j bedraagt, terwijl de waarden $y_{j+1}, \ldots, y_m$ op $\binom{f_j}{m-j}$ manieren kunnen worden gekozen. Immers, deze waarden moeten worden gekozen uit de verzameling $\{0, 1, \ldots, f_j - 1\}$, zonder teruglegging, waarna de volgorde, i.v.m. (12) meteen vastligt. Er geldt dus

$$N_j \binom{f_j}{m-j} = A_j + A_{j+1} . \qquad (13)$$

Uit (13) volgt nu

$$\sum_{j=0}^{m-1} (-1)^j N_j \binom{f_j}{m-j} = \sum_{j=0}^{m-1} (-1)^j (A_j + A_{j+1}) = (-1)^{m-1} A_m = (-1)^{m-1} N_m,$$

waarmee de stelling bewezen is.

NARAYANA & FULTON [17] geven een verwante formule voor het aantal der composities (geordende partities) van een bepaald getal die worden gedomineerd door een gegeven compositie van dat getal.

6.1 Voor een speciale keuze van f_j kan men in principe N_j bepalen m.b.v. (8). Een andere toepassingsmogelijkheid is het vinden van identiteiten door voor f_j een reeks te kiezen waarvan de bijbehorende N_j bekend is. Als bijv. $f_j = \min(j+1, b+1)$, dan volgt uit (3) met $m = 1$ en $a = b$ dat $N_j = \frac{1}{j+1}\binom{2j}{j}$, zodat

$$\sum_{j=0}^{r} \frac{(-1)^j}{j+1} \binom{2j}{j} \binom{j+1}{r-j} = 0 \qquad (r > 0).$$

6.2 Indien de reeks

$$\mathcal{N}(t) = \sum_{j=0}^{\infty} N_j t^j \qquad (14)$$

convergeert, kan men voor $j > 0$ N_j in (14) vervangen door

$$(-1)^{j-1} \sum_{i=0}^{j-1} (-1)^i N_i \binom{f_i}{j-i} .$$

Het blijkt dan dat $\mathcal{N}(t)$ er uit valt, en men vindt

$$\sum_{j=0}^{\infty} N_j t^j (1-t)^{f_j} = 1 . \qquad (15)$$

Deze relatie kan worden geïnterpreteerd door een stochastische wandeling vanuit (0,0) te beschouwen waarbij een stap naar rechts steeds een kans t heeft ($0 < t < 1$), en een stap omhoog een kans $1 - t$.

6.3 Neemt men in (15) voor f_j een rekenkundige reeks $\alpha j + \beta$ met α en β geheel, $\alpha \geq 0$, $\beta \geq 1$, dan vindt men

$$\sum_{j=0}^{\infty} N_j u^j = (1 - t)^{-\beta} \tag{16}$$

waarin t en u als volgt van elkaar afhangen

$$t(1 - t)^{\alpha} = u. \tag{17}$$

We merken hierbij op dat dit verband voor voldoend kleine waarden van t en u één-éénduidig is.

Voor $\beta = 1$, dus $f_j = \alpha j + 1$, geldt [1)]

$$N_j = \frac{1}{\alpha j + 1} \binom{\alpha j + j}{j}. \tag{18}$$

Immers, het aantal toegelaten paden naar $(j, \alpha j + 1)$ is gelijk aan het aantal paden van $(j, \alpha j)$ naar (0,0) met stappen omlaag en naar links, waarvan alle vertices op of onder de lijn $y = \alpha x$ liggen. Door een geschikte draaiing van het rooster ziet men dat dit aantal, volgens (3) gelijk is aan de door (18) gegeven waarde. Substitutie in (16) geeft dan

$$\sum_{j=0}^{\infty} \frac{1}{\alpha j + 1} \binom{\alpha j + j}{j} u^j = \frac{1}{1 - t} \tag{19}$$

of ook

$$\sum_{j=1}^{\infty} \frac{1}{j} \binom{\alpha j + j}{j - 1} u^j = \frac{1}{1 - t}, \tag{20}$$

een formule die ook voorkomt in PÓLYA en SZEGÖ [19] p. 301.

---

1) Oorspronkelijk ontleende ik (18) aan RUNNENBURG [18], waarin deze relatie wordt gegeven als illustratie van veel algemenere formules. Later bleek dat de in [18] gegeven afleiding enigszins kon worden verkort door gebruikmaking van (3).

6.4 Een variant op (8) is de volgende formule

$$N_m = \binom{h_m}{m} - \sum_{i=0}^{m-1} \binom{h_m - h_i}{m - i} N_i \tag{21}$$

waarin $h_m = f_m + m$. Het bewijs is zeer eenvoudig: het totale aantal paden van $(0,0)$ naar (m,f_m) is $\binom{h_m}{m}$. De niet-toegestane paden kan men onderscheiden naar het eerste punt van de vorm (i,f_i) met $0 \leq i \leq m-1$. Het aantal dezer paden is N_i maal $\binom{h_m - h_i}{m - i}$, enz. Substitueert men (21) in (14) dan vindt men voor het lineaire geval $f_j = \alpha j + \beta$ de formule

$$\mathcal{N}(t) = \frac{\sum_{j=0}^{\infty} \binom{\alpha j + j + \beta}{j} t^j}{\sum_{j=0}^{\infty} \binom{\alpha j + j}{j} t^j} . \tag{22}$$

Voor $\beta = 1$ kunnen we (18) weer toepassen:

$$\sum_{j=0}^{\infty} \frac{1}{\alpha j + 1} \binom{\alpha j + j}{j} t^j = \frac{\sum_{j=0}^{\infty} \binom{\alpha j + j + 1}{j} t^j}{\sum_{j=0}^{\infty} \binom{\alpha j + j}{j} t^j} . \tag{23}$$

LITERATUUR:

1. J. BERTRAND, Solution d'un problème; C.R. des Séances de l'Acad. Sci. Paris, 105 (1887), p. 369.
2. D. ANDRÉ, Solution directe du problème résolu par M. Bertrand; C.R. des Séances de l'Acad. Sci. Paris, 97 (1887), p. 436-437.
3. D.E. BARTON and C.L. MALLOWS, Some aspects of the random sequence: a review of amalgamation, ballots, records, sorting & Simon Newcomb's problem; Syllabus van een in 1963 te Kopenhagen gehouden lezing.
4. W.A. WHITWORTH, Choice and Chance; (4e druk) Deighton Bell, Cambridge 1886.
5. K.L. CHUNG and W. FELLER, Fluctuations in coin tossing; Proc. Nat. Acad. Sci. U.S.A. 35 (1949), p. 605-608.
6. J.L. HODGES, Galton's rank order test; Biometrika 42 (1955), p. 261-262.
7. O. ENGELBERG, Generalizations of the ballot problem; Rapport van de Columbia University, New York, sept. 1963.
8. F. GÖBEL, Some remarks on ballot problems; Rapport S 321a van het Mathematisch Centrum, Amsterdam, 1964.
9. L. TAKÁCS, The distribution of majority times in a ballot; Z. Wahrscheinlichkeitstheorie 2 (1963), p. 118-121.
10. L. TAKÁCS, Ballot problems; Z. Wahrscheinlichkeitstheorie 1 (1962), p. 154-158.
11. E. BARBIER, Generalisation du problème résolu par M. Bertrand; C.R. des Séances de l'Acad. des Sci. Paris, 105 (1887), p. 407.
12. A. AEPPLI, Zur Theorie verketteter Wahrscheinlichkeiten; Proefschrift, Zürich 1924.

13. A. DVORETZKY en T. MOTZKIN, A problem of arrangements, Duke Math. J. 14 (1947), p. 305-313.

14. H.D. GROSSMAN, Paths in a lattice triangle; Scripta Math. 16 (1950), p. 207-212.

15. M.T.L. BIZLEY, Derivation of a new formula for the number of minimal lattice paths from (0,0) to (km,kn) having just t contacts with the line my = nx and having no points above this line; and a proof of Grossman's formula for the number of paths which may touch but do not rise above this line; J. Inst. Actuar. 80 (1954), p. 55-62.

16. L. TAKÁCS, A generalization of the ballot problem and its application in the theory of queues; J. Amer. Statist. Assoc. 57 (1962), p. 327-337.

17. T.V. NARAYANA en G.E. FULTON, A note on the compositions of an integer; Canadian Math. Bull. 1 (1958), p. 169-173.

18. J. Th. RUNNENBURG, Random walk problemen; Rapport S 323 van het Mathematisch Centrum, Amsterdam, 1964.

19. G. PÓLYA en G. SZEGÖ, Aufgaben und Lehrsätze aus der Analysis I;(First American Edition, New York 1945).

typ: HdH