A királyfi párválasztása

Ruritániában hagyomány, hogy amikor a királyfi eléri a kiskorúságot, egyenként felvonultatják előtte az ország szűzeit, és ő választ egyet közülük. A királyfi annyit tud, hogy hány szűz van összesen, hogy nincs közöttük két egyforma szépségű, és hogy véletlen sorrendben mutatják be neki őket. A szabály: a királyfi vagy elfogadja az éppen bemutatott szűzet (és ezzel vége a ceremóniának), vagy elküldi azt és megszemléli a következőt (elküldött szűzet visszaszólítani nem lehet).

Mi a királyfi optimális stratégiája, ha minél szebb szűzet akar kapni? (Hogy *létezik* optimum, az nyilvánvaló: véges az eseménytér.)

A szűzek száma N; feltehetjük, hogy $N > 1$, különben nincs is feladat. A szűzek szépségrangja $1, \ldots, N$; nagyobb rang nagyobb szépség. (Nem tudjuk, hogy a szépségük mérhető-e, csak azt, hogy rendezhető.) Nevezzük a k rangú szűzet k-nak.

A rang normalizálásával rendeljünk a szűzekhez értéket úgy, hogy az átlagos értékük (N-től függetlenül) $1/2$ legyen: így a k szűz értéke $k/(N+1)$. (A rangsoron ez nem változtat, csak jobban összehasonlíthatóvá teszi a különböző N-ekhez tartozó választási procedúrákat.)

Tekintsük a következő stratégiát: A királyfi megnézi és elküldi az első n szűzet, megjegyezve annyit, hogy közülük milyen szép a legszebb — ez a mintavételi szakasz —, majd a továbbiak közül kiválasztja magának az első olyat, aki szebb a megjegyzettnél (vagy, ami ugyanaz, az összes addigiaknál), ha pedig ilyen nincs, akkor választja az utolsót — ez a döntési szakasz.

Ha $n = 0$, azaz rögtön az elsőt választja, akkor a várható érték $1/2$.

A továbbiakban legyen $n > 0$ (és persze $n < N$).

Ha az N szűz történetesen az első n között van, akkor az utolsó marad a választott, és a várható érték $N/2(N+1) = 1/2 - 1/2(N+1)$ (az N-től különbözőek átlaga). Ha az első n között a legszebb az $N - j$, ahol $j > 0$, akkor a királyfi talál majd szebbet: valamelyiket az $N, \ldots, N - j + 1$ közül, a várható érték így $(2N + 1 - j)/2(N + 1) = 1 - (j + 1)/2(N + 1)$ (az $N - j$-nél jobbak átlaga).

Jelölje $p_j^{(N,n)} = p_j$ annak az eseménynek a valószínűségét, hogy az első n között legszebbként ott van $N - j$:

$$p_j = \frac{n}{N-j} \cdot \frac{(N-n)(N-n-1)\cdots(N-n-j+1)}{N(N-1)\cdots(N-j+1)} = \frac{\binom{N-j-1}{n-1}}{\binom{N}{n}}$$

(belátható indukcióval). Ez $j = 0, \ldots, N - n$-re teljes eseményrendszer.

A várható érték tehát ($n > 0$):

$$\mathsf{M} = \frac{1}{\binom{N}{n}} \sum_{0}^{N-n} {}_{j} \left(1 - \frac{j+1}{2(N+1)}\right) \binom{N-j-1}{n-1} - \frac{n}{2N}$$

(az utolsó tag a korrekció a $j = 0$ esetben eltérő várható érték miatt).

Észrevéve, hogy $\binom{N+1}{n+1} = \sum_{0}^{N-n}{}_{j} \binom{N-j}{n} = \sum_{0}^{N-n}{}_{j} \frac{N+1-(j+1)}{n} \binom{N-j-1}{n-1}$,
M-et egyszerűbb alakra lehet hozni:

$$\mathsf{M} = 1 - \frac{1}{2(n+1)} - \frac{n}{2N}$$

(ez egyébként $n = 0$-ra is a helyes értéket adja).

M maximuma $n = \sqrt{N} - 1$-nél van (belátható deriválással);

$$\mathsf{M}_{\max} = 1 - \frac{1}{\sqrt{N}} + \frac{1}{2N}.$$

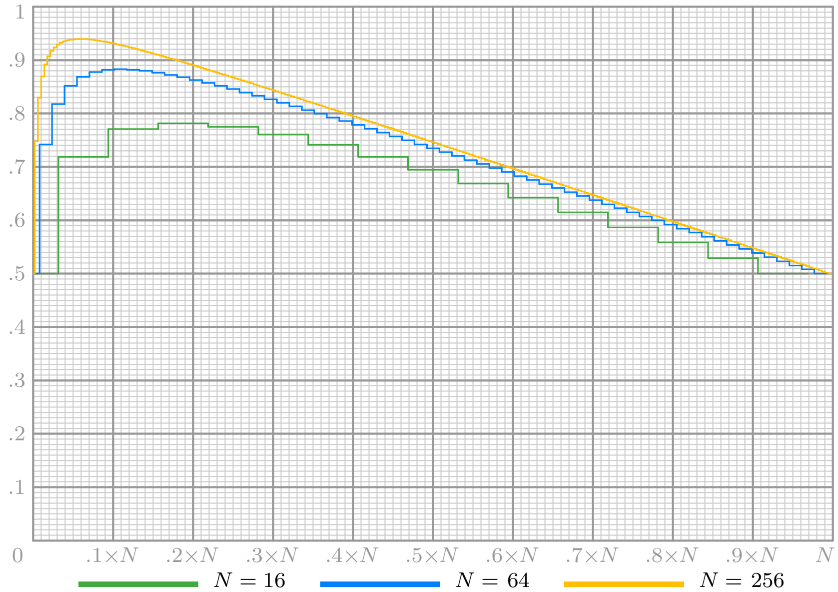
n-nek persze egész számnak kell lennie. Ha N teljes négyzet, legyen $\nu = \sqrt{N}$, különben pedig legyen ν a $\sqrt{N}$ két egészre-kerekítése közül valamelyik. Legyen $R(N,\nu) = (\sqrt{N} - \nu)^2 / 2N\nu$ és $R(N) = \min_{\nu} R(N,\nu)$.[1] Akkor

$$\mathsf{M}_{\mathrm{opt}} = 1 - \frac{1}{2\nu} - \frac{\nu - 1}{2N} = \mathsf{M}_{\max} - R(N).$$

$R(N) \le \frac{1}{4} N^{-2/3}$ (mert valamelyik ν-re $\left|\sqrt{N} - \nu\right| \le 1/2$).[2]

Vagyis ez a stratégia a királyfit várhatólag a legszebb $\sqrt{N}$ szűz közül egy csúnyácskábbhoz juttathatja.[3]

M grafikonja N néhány értékére



[1] Az $R(N)$-t minimalizáló kerekítés megegyezik a közönséges kerekítéssel, kivéve azt, hogy $N = \lfloor\sqrt{N}\rfloor \cdot \lceil\sqrt{N}\rceil$ esetén (nemcsak $\nu = \lfloor\sqrt{N}\rfloor$, hanem) $\nu = \lceil\sqrt{N}\rceil$ is minimalizál. Nem igazolom, mert (mint mindjárt látni fogjuk) ennek tisztázása nélkül is kapunk megfelelő becslést.

[2] $R(N)$ becslésében jelentősen javítható az együttható; ezt sem írom ide.

[3] $N = 2$-re még nem jobb a találomnál; növekvő N-re tökélyhez közelítő szűzzel kecsegtet.

Intuitíve is világos, ki is számolható, hogy a döntési szakaszban (a) nem érdemes a mintabeli legszebbnél csúnyábbat választani (ráérünk megvárni az utolsót), (b) nem érdemes az első szebb után tovább várakozni (ez ekvivalens lenne a mintavételi szakasz random meghosszabbításával).

Más megvariálási lehetőség nincsen.[1]

[Intranzigens királyfi]

Mi a királyfi optimális stratégiája, ha inkább szűz nélkül marad, de nem éri be olyannal, akinél látott már szebbet is?

Most a várható érték szummája 1-től indul, korrekciós tag pedig nincs:

$$\widetilde{\mathsf{M}} = 1 - \frac{1}{2(n+1)} - \left(1 - \frac{1}{2(N+1)}\right)\frac{n}{N}$$

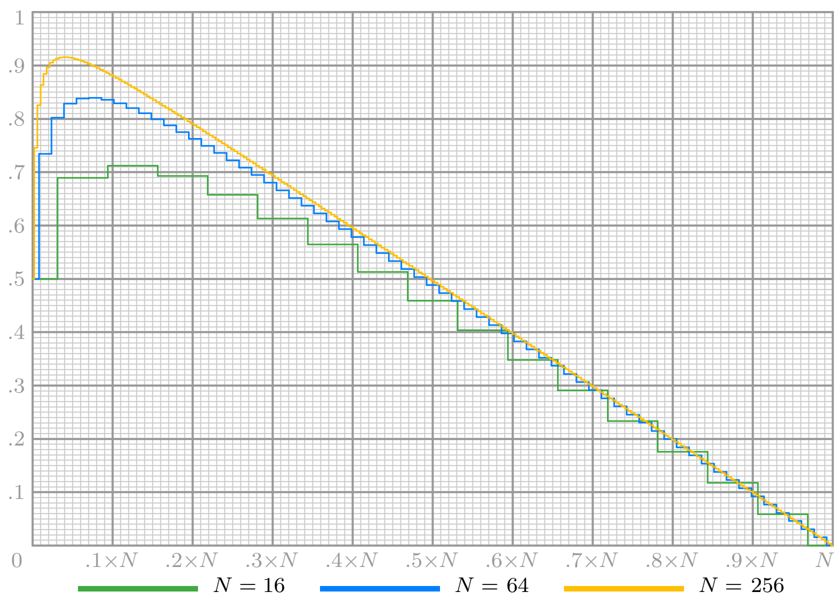
(ez $n = N$-re is helyes).

A maximum-hely: $n = \sqrt{\frac{N}{2}\left(1 + \frac{1}{2N+1}\right)} - 1$;

$$\widetilde{\mathsf{M}}_{\text{max}} = 1 - \sqrt{\frac{2}{N} + \frac{1}{N}}.$$

$$\widetilde{\mathsf{M}}_{\text{opt}} = \widetilde{\mathsf{M}}_{\text{max}} - \widetilde{R}; \quad \widetilde{R}(N) \ll N^{-3/2}.\text{[2]}$$

$\widetilde{\mathsf{M}}$ grafikonja N néhány értékére



[1] Ha a királyfi nemcsak szépség szerinti rendezést tud, hanem rendelkezik „szépségmértékkel” és ismeri a „szépségeloszlást” is, akkor a konkrétan kapott mintaelemek értékét felhasználhatja azokban a döntésekben, hogy kér-e további mintaelemet, és hogy a továbbiakban hogyan választ.

[2] $\widetilde{R}(N)$-t is ki lehet részletezni; nem írom ide.

[Maximalista királyfi]

Mi a királyfi optimális stratégiája, ha csakis a legszebb szűz kell neki?

Kedvező kimenetelek: Ha a legszebb szűz j-edikként kerül sorra, akkor ahhoz, hogy ő legyen a választott, szükséges és elegendő, hogy az előző $j-1$ szűz legszebbike az első n között (a mintában) legyen. Adódik:

$$n = 0\colon\ \bar{\mathsf{M}} = \frac{1}{N}, \qquad n > 0\colon\ \bar{\mathsf{M}} = \frac{n}{N}\sum_{n}^{N-1}{}_{j}\,\frac{1}{j}.$$ [1]

A deriválás most nem megy. Kis N-re $\bar{\mathsf{M}}$ numerikusan kiszámolható; nagy N-re, ha n is vele növekszik, $\bar{\mathsf{M}} \sim \frac{n}{N}\big(\log(N-1) - \log(n-1)\big) \sim \frac{n}{N}\log\frac{N}{n}$. Ha az $\frac{n}{N}$ arány tart valamilyen q-hoz, akkor

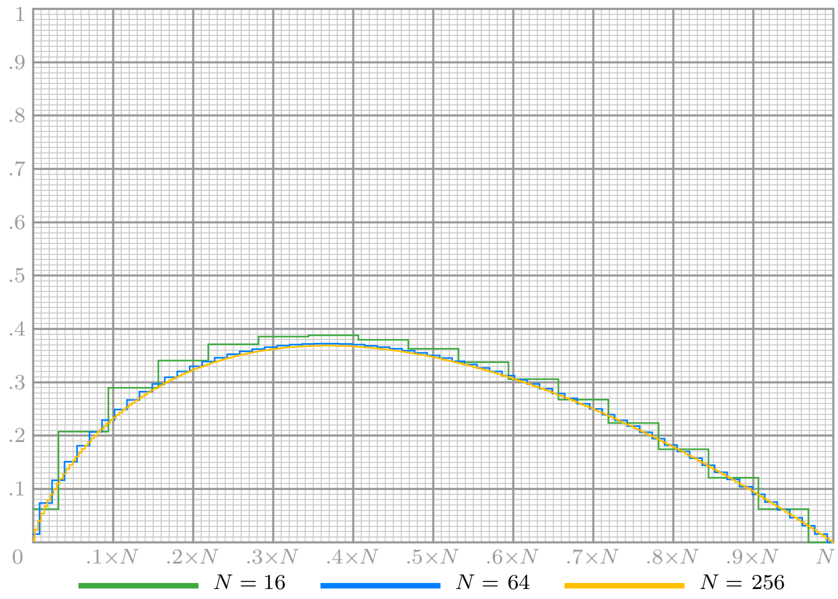
$$\bar{\mathsf{M}} \sim -q\log q.$$

Ez már deriválható; az optimum-hely e^{-1}, a függvény ott Ne^{-1}.

Ebből $\bar{\mathsf{M}}$ optimum-helyének közelítése Ne^{-1}; $\bar{\mathsf{M}}$ optimum-értékéé e^{-1}. (Feltételeztük az „aránytartást”, de visszatekintve nyilvánvaló, hogy ellenesetben bizonyosan nem kapnánk optimumot.)

A $-q\log q$ kis N-re sem rossz; az eltérések kiszámolhatók. Nagy N-re az eltérések becsülhetők a harmonikus sor és a logaritmus különbségének becsléséből.[2]

$\bar{\mathsf{M}}$ grafikonja N néhány értékére



---

[1] Vagy másképp: $n \in [0, N-1]$: $\bar{\mathsf{M}} = \frac{1}{N} + \frac{n}{N}\sum_{n+1}^{N-1}{}_{j}\,\frac{1}{j}$.

[2] Sok számolás; nem írom ide. A maximum-hely eltérése minden N-re < 1, de a közönséges kerekítés nem mindig jó — nem találtam szabályát, hogy mikor merre kell kerekíteni.

Érdekes: A megoldás jellegén az előző variáns kívánalom-szigorítása nem sokat változtatott, ez a mostani továbbszigorítás azonban alapvetően megváltoztatta.

Szintén érdekes (bár itt semmi haszna):

$$\binom{N}{n}^{-1} \sum_{n}^{N-1}{}_{j} \frac{\binom{j-1}{n-1}}{N-j} = \frac{n}{N} \sum_{n}^{N-1}{}_{j} \frac{1}{j}, \quad \text{ha } n \in [1, N-1].$$

A baloldalt úgy kapjuk, hogy a minták legszebb szűze szerint csoportosítva számláljuk össze a kedvező kimeneteleket. (A sorok formális manipulálásával nehéznek tűnik igazolni az egyenlőséget.)

[Melléktermék]

Mi a helyzet, ha szűzet vissza lehet hívni, a boltba vissza lehet menni, stb.? Ekkor persze nincs optimum-probléma; a probléma ilyenkor a *részleges információ alapján történő döntés*: az, hogy egy adott várható jóság eléréséhez hány példányt kell megvizsgálni, ill. (vice versa) hány példány alapján menynyire jó döntés várható.

Nincs külön mintavételi és döntési szakasz; n 1-től indul; az összefüggés most

$$\widehat{\mathsf{M}} = 1 - \frac{1}{(n+1)}.$$

Talán legegyszerűbb így belátni: Visszacsinálva a normalizálást (amelyben a k-adik érték $k/(N+1)$ volt), vagyis felszorozva $N+1$-gyel, a feladat így fogalmazható át: $1, 2, \ldots, N$ közül kiválasztunk n darabot — várhatólag mekkora lesz ezek közül a legnagyobb?

Annak valószínűsége, hogy r a legnagyobb, $n \le r \le N$:

$$p_r = \frac{r(r-1)\cdots(r-n+1) - (r-1)(r-2)\cdots(r-n)}{N(N-1)\cdots(N-n+1)} = \frac{\binom{r-1}{n-1}}{\binom{N}{n}};$$

a keresett várható érték tehát

$$\sum_{n}^{N}{}_{r}\, r p_r = \frac{n}{\binom{N}{n}} \sum_{n}^{N}{}_{r} \binom{r}{n} = (N+1)\frac{n}{n+1},$$

ahogy kell.

Hasonlóan (vagy szimmetria okából) a legkisebbnek a várható értéke $\dfrac{N+1}{n+1}$.

Szuggesztív speciális eset — négyzetgyök-ökölszabály —: Ha $\sqrt{N}$ elemű mintát veszünk, várhatólag benne lesz a legjobb $\sqrt{N}$ elem valamelyike (közülük egy rosszabbféle). Fordítva is elmondhatjuk: Ahhoz, hogy várhatólag ilyen jó elemet találjunk, elég ekkora mintát venni.