Variance du nombre de maxima dans les hypercubes

Christian Costermans

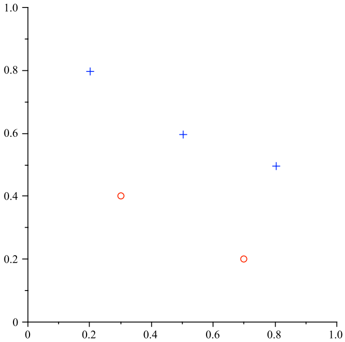
Université Lille 2

École Jeunes Chercheurs en Informatique Mathématique
19-23 Mars 2007

Plan

Problème initial

- On considère un ensemble de n points, v.a.i.i.d
- Répartition uniforme sur $[0,1]^d$, muni de l'ordre-produit.
- On note $K_{n,d}$ le nombre de maxima.



Un exemple avec $n = 5$ et $d = 2$.

Résultats antérieurs

- Barndorff-Nielsen et Sobel(1966), détermination de l'espérance

$$\mu_{n,d} = \mathbb{E}(K_{n,d}) = \sum_{n \geq n_1 \geq \ldots \geq n_{d-1} > 0} \frac{1}{n_1 \ldots n_{d-1}}.$$

Ecriture 2

  Détermination de la *distribution* pour {d petit, n quelconque} et {n petit, d quelconque}
- Ivanin(1976), détermination du moment d'ordre 2

$$\mathbb{E}(K_{n,d}^2) = \mu_{n,d} + \sum_{1 \leq t \leq d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \sum^{(*)} \frac{1}{i_1 \ldots i_{d-2} j_1 \ldots j_{d-1}},$$

  où la somme $(*)$ est prise sur tous les indices vérifiant $1 \leq i_1 \ldots \leq i_{t-1} \leq l, 1 \leq i_t \leq \ldots \leq i_{d-2} \leq l$ et $l+1 \leq j_1 \leq \ldots \leq j_{d-1} \leq n$.

Avancées récentes

$$\mu_{n,d} \sim \frac{1}{(d-1)!} \ln^{d-1}(n), \text{ mais}$$

- Devroye(1998), $Var(K_{n,d}) = \mathrm{O}(\mu_{n,d})$.
- Bai et al.(1998)

$$\begin{aligned} Var(K_{n,d}) &\sim \left(\frac{1}{(d-1)!} + \kappa_d\right) \ln^{d-1}(n), \\ \text{avec} \quad \kappa_d &= \sum_{t=1}^{d-2} \frac{1}{t!(d-1-t)!} \sum_{l \geq 1} \frac{1}{l^2} \sum^{(**)} \frac{1}{i_1 \ldots i_{t-1} j_1 \ldots j_{d-2-t}} \end{aligned}$$

la somme $(**)$ calculée sur tous les indices vérifiant $1 \leq i_1 \leq \ldots \leq i_{t-1} \leq l$ et $1 \leq j_1 \leq \ldots \leq j_{d-2-t} \leq l$.

Avancées récentes

Théorème (Bai et al., 2004)

$$\begin{aligned} Var(K_{n,d}) &= \sum_{0 \le j \le d-1} \frac{1}{(d-1-j)!} \left(\frac{(-1)^j}{j!} \Gamma^{(j)}(1) + c_{dj} \right) \ln^{d-1-j}(n) \\ &+ \mathrm{O}(n^{-1} \ln^{2d-2}(n)), \end{aligned}$$

où $c_{d,j}$ est défini comme le coefficient en u^j dans le développement de Taylor de

$$-\frac{2\Gamma(2-u)}{(d-1)!} \int_0^1 \frac{(-\ln x)^{d-1}}{(1+x)^{2-u}} dx + \sum_{1 \le k \le d-1} \frac{\binom{d}{k}\Gamma(2-u)}{(k-1)!(d-1-k)!}$$

$$\times \int_0^1 \int_0^1 \left(\frac{(-\ln x)^{k-1}(-\ln x)^{d-1-k}}{(x+z-xz)^{2-u}} - \frac{(-\ln x)^{k-1}(-\ln x)^{d-1-k}}{(x+z)^{2-u}} \right) dxdz.$$

Problème en suspens

" It remains open how the integrals in (1) may be further simplified. "

Objectif

traiter à l'aide d'outils combinatoires le calcul (exact, puis asymptotique) de cette variance.

Plan

Sommes harmoniques et polyzêtas

Nombres harmoniques généralisés

$$\mathrm{A}_r(N) = \sum_{n=1}^{N} \frac{1}{n^r} \qquad N \in \mathbb{N}, r \geq 0$$

Extension à des multi-indices $\underline{\mathbf{s}} = (s_1, \ldots, s_r)$

$$\mathrm{A}_{\underline{\mathbf{s}}}(N) = \sum_{N \geq n_1 \geq \ldots \geq n_r > 0} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}$$

Pour $s_1 > 1$, polyzêta large

$$\underline{\zeta}(\underline{\mathbf{s}}) = \lim_{N \to +\infty} \mathrm{A}_{\underline{\mathbf{s}}}(N) = \sum_{n_1 \geq \ldots \geq n_r > 0} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}$$

Polyzêtas large ou stricts

Soit $\underline{\mathbf{s}} = (s_1, \ldots, s_r)$. Pour $s_1 > 1$, polyzêta défini par

$$\zeta(\underline{\mathbf{s}}) = \sum_{n_1 > \ldots > n_r > 0} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}$$

Passage d'une valeur à l'autre :

$$\begin{aligned} \underline{\zeta}(2,1,1) &= \zeta(2,1,1) + \zeta(2,2) + \zeta(3,1) + \zeta(4) \\ \zeta(2,1,1) &= \underline{\zeta}(2,1,1) - \underline{\zeta}(2,2) - \underline{\zeta}(3,1) + \underline{\zeta}(4) \end{aligned}$$

Codage symbolique

Nous adoptons le codage suivant :

$$\underline{\mathbf{s}} = (s_1, \ldots, s_r) \longleftrightarrow w = y_{s_1} \cdots y_{s_r} \in Y^*,$$

$$\text{où } Y = \{y_i | i \in \mathbb{N} \setminus \{0\}\}.$$

Dorénavant, nous noterons $\mathrm{A}_{\underline{\mathbf{s}}}(N) = \mathrm{A}_w(N)$ et, pour $s_1 > 1$,

$$\lim_{N \to +\infty} \mathrm{A}_{\underline{\mathbf{s}}}(N) = \underline{\zeta}(\underline{\mathbf{s}}) = \underline{\zeta}(w).$$

Produit harmonique ou "stuffle"

$$\sum_{n=1}^{N}\frac{1}{n^{s_1}}\sum_{m=1}^{N}\frac{1}{m^{s_2}}=\sum_{N\geq n\geq m>0}\frac{1}{n^{s_1}m^{s_2}}+\sum_{N\geq m\geq n>0}\frac{1}{m^{s_2}n^{s_1}}-\sum_{N\geq n>0}\frac{1}{n^{s_1+s_2}}$$

$$\Rightarrow \mathrm{A}_{y_{s_1}}(N)\mathrm{A}_{y_{s_2}}(N)=\mathrm{A}_{y_{s_1}y_{s_2}}(N)+\mathrm{A}_{y_{s_2}y_{s_1}}(N)-\mathrm{A}_{y_{s_1+s_2}}(N)$$

Le *stuffle* de deux mots est défini par

$$\begin{aligned}\epsilon \uplus u &= u,\\ (y_i u)\uplus(y_j v) &= y_i(u\uplus y_j v)+y_j(y_i u\uplus v)-y_{i+j}(u\uplus v).\end{aligned}$$

Théorème

Pour tous $u, v\in Y^*$, $\mathrm{A}_u(N)\,\mathrm{A}_v(N)=\mathrm{A}_{u\uplus v}(N)$.

Exemple

$$\begin{aligned} y_2y_3 \sqcup y_1 &= y_2(y_3 \sqcup y_1) + y_1(y_2y_3 \sqcup \epsilon) - y_3(y_3 \sqcup \epsilon) \\ &= y_2(y_3y_1 + y_1y_3 - y_4) + y_1y_2y_3 - y_3^2 \\ &= y_2y_3y_1 + y_2y_1y_3 - y_2y_4 + y_1y_2y_3 - y_3^2. \end{aligned}$$

Donc

$$\begin{aligned} \mathrm{A}_{y_2y_3}(N)\mathrm{A}_{y_1}(N) &= \mathrm{A}_{y_2y_3y_1}(N) + \mathrm{A}_{y_2y_1y_3}(N) \\ &- \mathrm{A}_{y_2y_4}(N) + \mathrm{A}_{y_1y_2y_3}(N) - \mathrm{A}_{y_3^2}(N), \end{aligned}$$

ou encore

$$\begin{aligned} \mathrm{A}_{2,3}(N)\mathrm{A}_1(N) &= \mathrm{A}_{2,3,1}(N) + \mathrm{A}_{2,1,3}(N) \\ &- \mathrm{A}_{2,4}(N) + \mathrm{A}_{1,2,3}(N) - \mathrm{A}_{3,3}(N). \end{aligned}$$

Plan

Réecriture des deux premiers moments

Ecriture1

$$\mu_{n,d} = \mathrm{A}_{y_1^{d-1}}(n),$$

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{t=1}^{d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

avec, pour $w = y_{s_1} \ldots y_{s_r}$,

$$\mathrm{A}_w(n; l) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq l} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}.$$

Réecriture des deux premiers moments

Ecriture1

$$\mu_{n,d} = \mathrm{A}_{y_1^{d-1}}(n),$$

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{t=1}^{d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

avec, pour $w = y_{s_1} \dots y_{s_r}$,

$$\mathrm{A}_w(n; l) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq l} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}.$$

Réecriture des deux premiers moments

Ecriture1

$$\mu_{n,d} = \mathrm{A}_{y_1^{d-1}}(n),$$

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{t=1}^{d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

avec, pour $w = y_{s_1} \ldots y_{s_r}$,

$$\mathrm{A}_w(n; l) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq l} \frac{1}{n_1^{s_1} \ldots n_r^{s_r}}.$$

Réecriture des deux premiers moments

Ecriture1

$$\mu_{n,d} = \mathrm{A}_{y_1^{d-1}}(n),$$

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{t=1}^{d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

avec, pour $w = y_{s_1} \dots y_{s_r}$,

$$\mathrm{A}_w(n;l) = \sum_{N \geq n_1 \geq \ldots \geq n_r \geq l} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}.$$

Réecriture des deux premiers moments

Ecriture1

$$\mu_{n,d} = \mathrm{A}_{y_1^{d-1}}(n),$$

$$\mathbb{E}(K_{n,d}^2) = \mathrm{A}_{y_1^{d-1}}(n) + \sum_{t=1}^{d-1} \binom{d}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{d-t-1}}(l) \mathrm{A}_{y_1^{d-1}}(n; l+1),$$

avec, pour $w = y_{s_1} \dots y_{s_r}$,

$$\mathrm{A}_w(n; l) = \sum_{n \geq n_1 \geq \ldots \geq n_r \geq l} \frac{1}{n_1^{s_1} \dots n_r^{s_r}}.$$

Les détails techniques

Théorème

$\mathrm{A}_{y_1^r}$ s'exprime comme une combinaison polynômiale des A_{y_k}, pour $1 \le k \le r$

Exemple

$$\begin{aligned} \mathrm{A}_{y_1^2} &= \frac{1}{2}\left(\mathrm{A}_{y_1}^2 + \mathrm{A}_{y_2}\right) \\ \mathrm{A}_{y_1^3} &= \frac{1}{6}\mathrm{A}_{y_1}^3 + \frac{1}{2}\mathrm{A}_{y_1}\mathrm{A}_{y_2} + \frac{1}{3}\mathrm{A}_{y_3}. \end{aligned}$$

Proposition

Pour tout $w \in Y^*$, et $s > 0$

$$\sum_{l=1}^{n} \frac{1}{l^s}\mathrm{A}_w(l) = \mathrm{A}_{y_s w}(n).$$

En dimension 3

$$\begin{aligned}
\mathbb{E}(K_{n,3}^2) &= \mathrm{A}_{y_1^2}(n) + \sum_{t=1}^{2} \binom{3}{t} \sum_{l=1}^{n-1} \frac{1}{l} \mathrm{A}_{y_1^{t-1}}(l) \mathrm{A}_{y_1^{2-t}}(l) \mathrm{A}_{y_1^2}(n; l+1) \\
&= \mathrm{A}_{y_1^2}(n) + \sum_{l=1}^{n-1} \frac{6}{l} A_1(l) \Bigg(\frac{(\mathrm{A}_1(n) - \mathrm{A}_1(l))^2}{2} \\
&+ \frac{\mathrm{A}_2(n) - \mathrm{A}_2(l)}{2} \Bigg)
\end{aligned}$$

$$\begin{aligned}
Var(K_{n,3}) &= \mathrm{A}_{y_1^2}(n) + 6\mathrm{A}_{y_1^2}(n)\mathrm{A}_{y_1^2}(n-1) - 12\mathrm{A}_1(n)\mathrm{A}_{y_1^3}(n-1) \\
&+ 6\mathrm{A}_1(n) A_{y_1 y_2}(n-1) + 18\mathrm{A}_{y_1^4}(n-1) - 12\mathrm{A}_{y_1^2 y_2}(n-1) \\
&- 12\mathrm{A}_{y_1 y_2 y_1}(n-1) + 6\mathrm{A}_{y_1 y_3}(n-1) - \mathrm{A}_{y_1^2}^2(n).
\end{aligned}$$

Passage à l'asymptotique

Soit $\mathcal{Z}$ la $\mathbb{Q}$-algèbra engendrée par les polyzêtas $\{\zeta(w)\}_{w\in Y^*\setminus y_1Y^*}$ et la constante d'Euler γ.

Théorème

Pour tout $w\in Y^*$, il existe des coefficients $b_i\in\mathcal{Z}$ tels que

$$\mathrm{A}_w(N)\sim\sum_{i=0}^{+\infty}b_iN^{\eta_i}\log^{\kappa_i}(N),\quad quand\quad N\rightarrow+\infty,$$

avec $\kappa_i\in\mathbb{N}$ et $\eta_i\in\mathrm{Z}$.

En conclusion

$$
\begin{aligned}
\mathbb{V}ar(K_{n,5}) &= \left(\frac{1}{4!}+\frac{33}{40}\zeta(2)^2\right)\ln^4(n) + \left(\frac{1}{6}\gamma - \frac{98}{3}\zeta(5) + \frac{33}{10}\zeta(2)^2\gamma - \frac{13}{3}\zeta(2)\zeta(3)\right)\ln^3(n) \\
&+ \left(\frac{10123}{140}\zeta(2)^3 + \frac{47}{2}\zeta(3)^2 + \frac{99}{20}\zeta(2)^2\gamma^2 + \frac{1}{4}\gamma^2 + \frac{1}{4}\zeta(2) - 13\zeta(2)\zeta(3)\gamma \right. \\
&- \left. 98\zeta(5)\gamma\right)\ln^2(n) + \left(\frac{1}{6}\gamma^3 + \frac{33}{10}\zeta(2)^2\gamma^3 + \frac{1}{2}\zeta(2)\gamma - 950\zeta(7)\right. \\
&- 13\zeta(2)\zeta(3)\gamma^2 + 47\zeta(3)^2\gamma + \frac{1}{3}\zeta(3) - \frac{317}{5}\zeta(3)\zeta(2)^2 + \frac{10123}{70}\zeta(2)^3\gamma \\
&- \left. 98\zeta(5)\gamma^2 - 222\zeta(2)\zeta(5)\right)\ln(n) - \frac{13}{3}\zeta(2)\zeta(3)\gamma^3 + \frac{47}{2}\zeta(3)^2\gamma^2 \\
&- \frac{317}{5}\zeta(3)\zeta(2)^2\gamma - \frac{98}{3}\zeta(5)\gamma^3 + \frac{33}{40}\zeta(2)^2\gamma^4 + \frac{32}{3}\zeta(3)\zeta(5) + \frac{10123}{140}\zeta(2)^3\gamma^2 \\
&- 222\zeta(2)\zeta(5)\gamma + \frac{1}{24}\gamma^4 - 950\zeta(7)\gamma + 50\zeta(6,2) + \frac{1}{4}\zeta(2)\gamma^2 + \frac{1}{3}\zeta(3)\gamma \\
&+ \frac{9}{40}\zeta(2)^2 + \frac{95}{6}\zeta(2)\zeta(3)^2 + \frac{134739}{350}\zeta(2)^4 + o(1)
\end{aligned}
$$

Merci de votre attention.