Sampling algorithms related to determinantal processes

Mirjana Vuletić

University of Massachusetts Boston

November 2018

General picture:

- universal distributions (universal scaling limits)
- determinantal processes
- sampling algorithms
- probability, combinatorics, representation theory, statistical physics, number theory, algebraic geometry, tropical geometry, machine learning

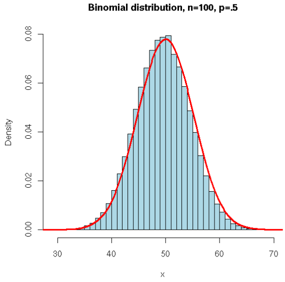
Applications (two examples):

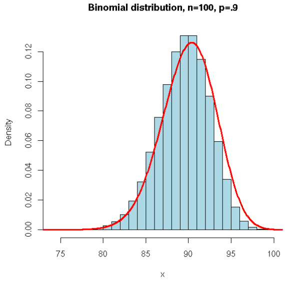
- combinatorics, limit shapes
- machine learning applications (Kulesza and Taskar [2012])

Sampling algorithms:

- Schur processes (joint work with Betea et al. [2014])
- determinantal processes on finite spaces (due to Hough et al. [2006])

Central limit theorem





Let $X_1, X_2, \ldots$ be independent Bernoulli trials $X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$

$$S_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n,p)$$

$$\frac{S_n - nE(X)}{SD(X)/\sqrt{n}} \to \mathcal{N}(0,1)$$

Limit theory for random matrices

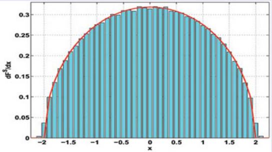
Wigner matrices

M is Hermitian ($M = \overline{M}^T$) and M_{ij} for $i \leq j$ are independent such that $E(M_{ij}) = 0$, $Var(M_{ii}) = \sigma^2$ and $Var(M_{ij}) = 1$, $i \neq j$

Semicircle law

Let M_n be a Wigner matrix of size n.

$$\frac{\#\text{of eigenvalues of} \frac{1}{\sqrt{n}} M_n}{n} \rightarrow$$



Universality: Erdös-Schlein-Yau and Tao-Vu [2010]

Asymptotics of any eigenvalue statistic $f(\lambda_1, \ldots, \lambda_n)$ of Wigner matrices is independent of the underlying distribution.

Special case of Wigner matrices:

GUE- Gaussian Unitary Ensemble

- M_{ij} are complex $\mathcal{N}(0,1)$ for $i \neq j$
- M_{ii} are real $\mathcal{N}(0,1)$

- GUE is a determinantal process; the joint density of eigenvalues can be written as a determinant
- eigenvalue spacing is the same as the spacing of zeros of Riemann zeta function:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

- The Riemann hypothesis: Riemann zeta function has its zeros only at the negative even integers and complex numbers with real part $1/2$.

Determinantal point processes

- Let $\mathcal{X} = \{x_1, \ldots, x_N\}$ be a finite space.
- A random point process P on $\mathcal{X}$ is a probability measure on the set $2^{\mathcal{X}}$ of all subsets of $\mathcal{X}$.
- Correlation function at $(x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ is defined as

$$\rho(x_1, x_2, \ldots, x_n) = \Pr(X | \{x_1, x_2, \ldots, x_n\} \subset X).$$

- P is determinantal if there exists a kernel $K(x, y)$ such that

$$\rho(x_1, x_2, \ldots, x_n) = \det(K(x_i, x_j))_{i,j=1\ldots n}.$$

- introduced by Macchi [1975], term "determinantal" was first used by Borodin-Olshanski [2000]

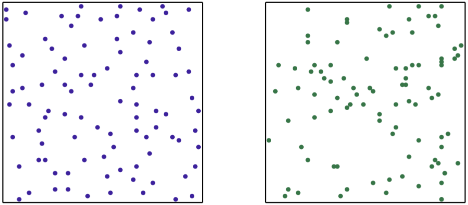
- determinantal processes exhibit repulsion (negative correlation) if we assume K is symmetric

$$\rho(x) = |K(x,x)| = K(x,x)$$

$$\rho(x_i, x_j) = \begin{vmatrix} K(x_i, x_i) & K(x_i, x_j) \\ K(x_j, x_i) & K(x_j, x_j) \end{vmatrix} = \rho(x_i)\rho(x_j) - K(x_i, x_j)^2$$

$$\rho(xi, x_j) = \rho(x_i)\rho(x_j|x_i) \leq \rho(x_i)\rho(x_j)$$

$$\rho(x_j|x_i) \leq \rho(x_j)$$



DPP Independent

General picture

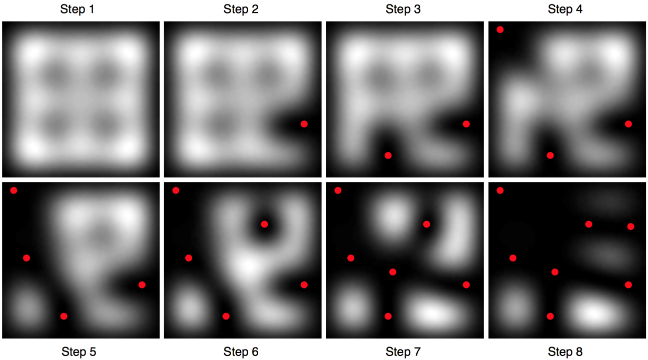
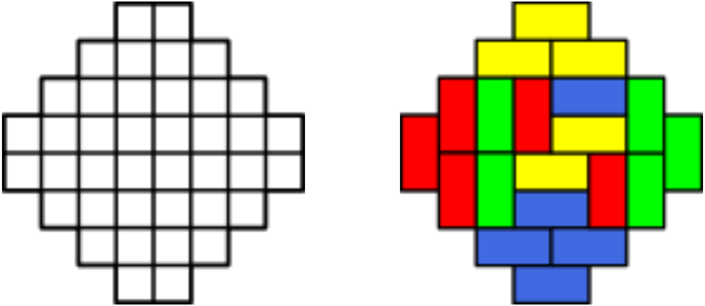


Figure taken from Kulesza and Taskar (20012)

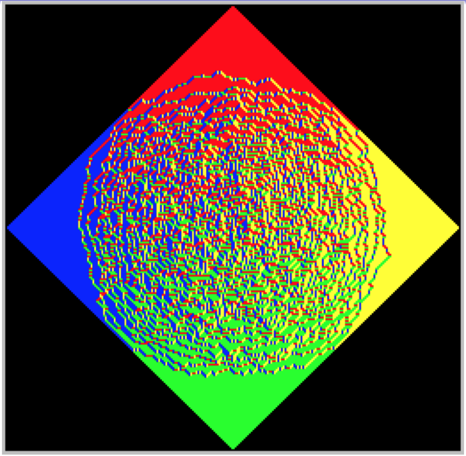
An example of a sampling algorithm (for domino tilings of an Aztec diamond)

Simulation (Patrik Ferrari)



- Aztec diamond of size n- four staircases with n steps glued together
- There are $2^{n(n+1)/2}$ different domino tilings of Aztec diamond of size n. [Elkies, Kuperberg, Larsen, Propp '92]

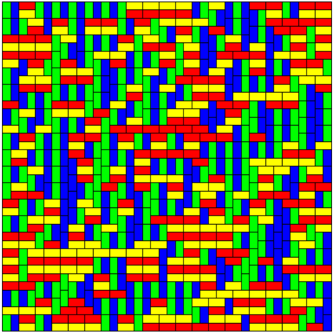
Arctic Circle Theorem



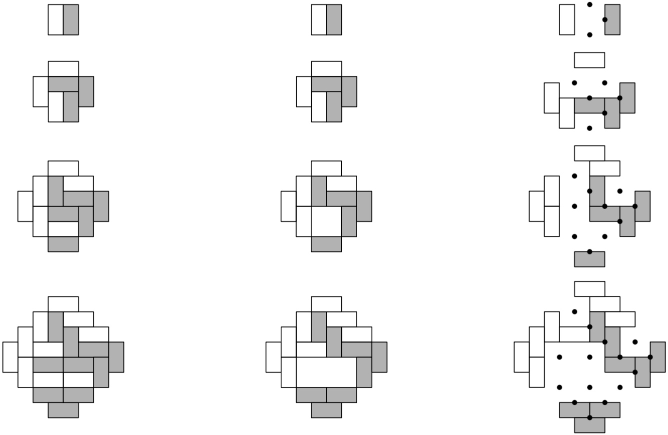
Arctic Circle Theorem [Jockusch, Propp, Shor 1985]

Domino tilings of a large Aztec diamond tend to be frozen outside the circle.

What about a square?



Shuffling algorithm

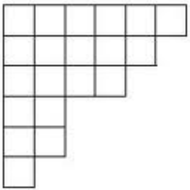


Algorithm due to Elkies, Kuperberg, Larsen, Propp (1992)

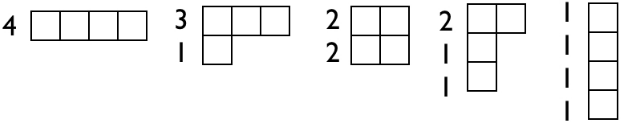
Figure taken from Nordenstam (2008)

Partitions

$\lambda = (6, 5, 4, 2, 2, 1)$



- size $|\lambda|$- sum of all entries (or the area)
- partitions of size 4:

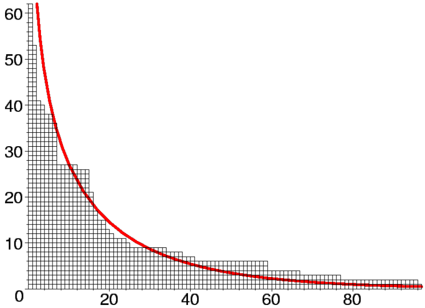


Limit shape- partitions

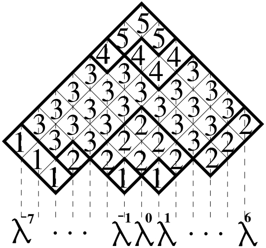
Vershik-Kerov [1985]

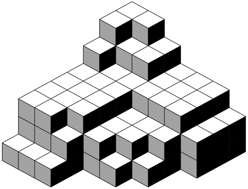
Limit shape of uniformly distributed paritions of size n, when $n \to \infty$ is

$$e^{-x} + e^{-y} = 1$$



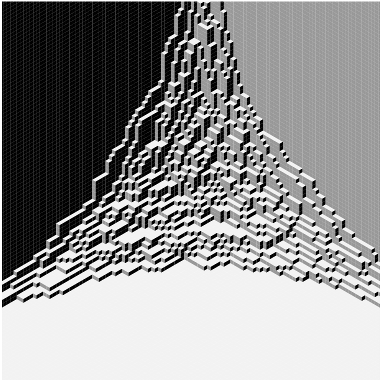
Plane partitions

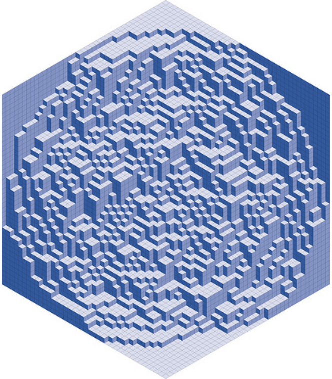




Limit shape- plane partitions

- uniform distribution on partitions of volume n, when $n \to \infty$: Cerf-Kenyon [01] and Okounkov-Reshetikhin [03]
- uniform distribution on partitions that fit in a box of size $n \times n \times n$, when $n \to \infty$: Cohn-Larsen-Propp [98]





Machine learning- major threads from newspaper corpus

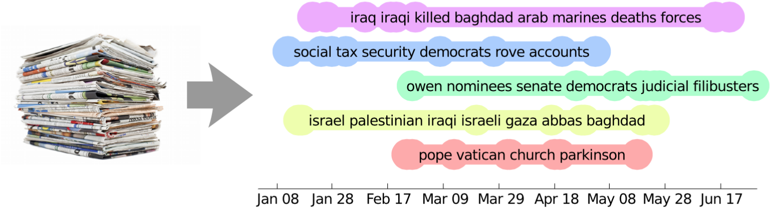


Figure taken from Kulesza and Taskar (20012)

Machine learning- major events in a thread

Thread: **pope vatican church parkinson**

Feb 24:	Parkinson's Disease Increases Risks to Pope
Feb 26:	Pope's Health Raises Questions About His Ability to Lead
Mar 13:	Pope Returns Home After 18 Days at Hospital
Apr 01:	Pope's Condition Worsens as World Prepares for End of Papacy
Apr 02:	Pope, Though Gravely Ill, Utters Thanks for Prayers
Apr 18:	Europeans Fast Falling Away from Church
Apr 20:	In Developing World, Choice [of Pope] Met with Skepticism
May 18:	Pope Sends Message with Choice of Name

Figure taken from Kulesza and Taskar (20012)

Machine learning- human pose detection

The output of a human pose detector is noisy and uncertain; applying diversity as a filter leads to a clean, separated set of predictions.

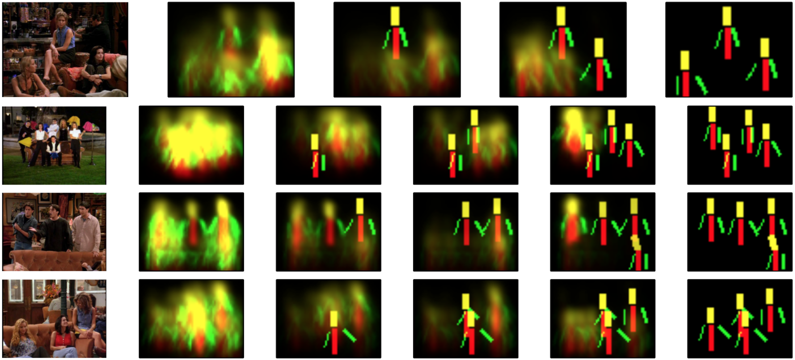
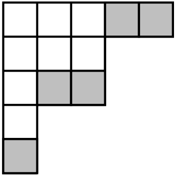


Figure taken from Kulesza and Taskar (20012)

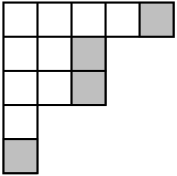
Interlacing

- interlacing $\lambda \succ \mu : \lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \cdots$
- dual interlacing $\lambda \succ' \mu$ means $\lambda' \succ \mu'$



$(5,3,3,1,1) \succ (3,3,1,1)$

horizontal strip



$(5,3,3,1,1) \succ' (4,2,2,1)$

vertical strip

- $w = (w_1, w_2, \ldots, w_n) \in \{\prec, \succ, \prec', \succ'\}^n$: w-interlaced sequences of partitions $\Lambda = (\emptyset = \lambda(0), \lambda(1), \ldots, \lambda(n) = \emptyset)$ means $\lambda(i-1) w_i \lambda(i)$, $\forall i$

Sampling of the Schur process

For a word $w = (w_1, w_2, \ldots, w_n) \in \{\prec, \succ, \prec', \succ'\}^n$, the *Schur process* of word w with parameters $Z = (z_1, \ldots, z_n)$ is a measure on the set of w-interlaced sequences of partitions $\Lambda = (\emptyset = \lambda(0), \lambda(1), \ldots, \lambda(n) = \emptyset)$ given by

$$Prob(\Lambda) \propto \prod_{i=1}^{n} z_i^{||\lambda(i)|-|\lambda(i-1)||}.$$

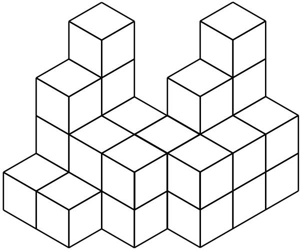
Remark.

$$s_{\lambda/\mu}(x_1) = x_1^{|\lambda|-|\mu|}\delta_{\lambda\succ\mu}.$$

Reverse plane partitions

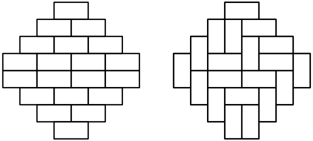
$(m \times n)$-boxed *plane partitions*: $w = (\prec)^m(\succ)^n$

1	3	4		
1	2	2		
0	2	2	3	4
0	0	2	2	2



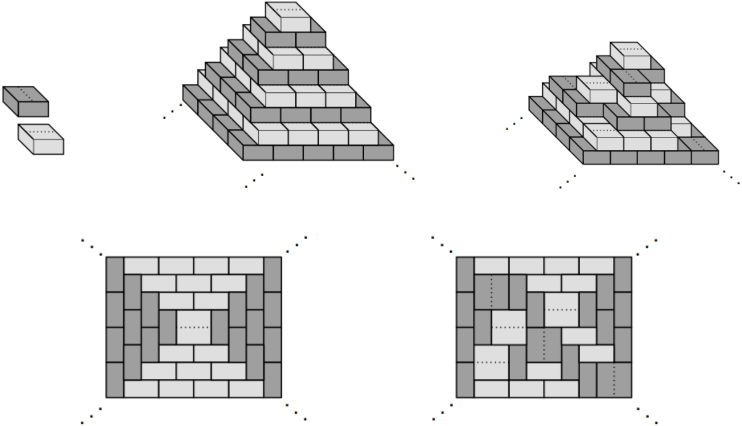
$$\emptyset \prec (1) \prec (3,1) \prec (4,2) \succ (2,2) \succ (2) \prec (3,2) \prec (4,2) \succ (2) \succ \emptyset$$

Aztec diamonds



$$w = (\prec', \succ)^n$$

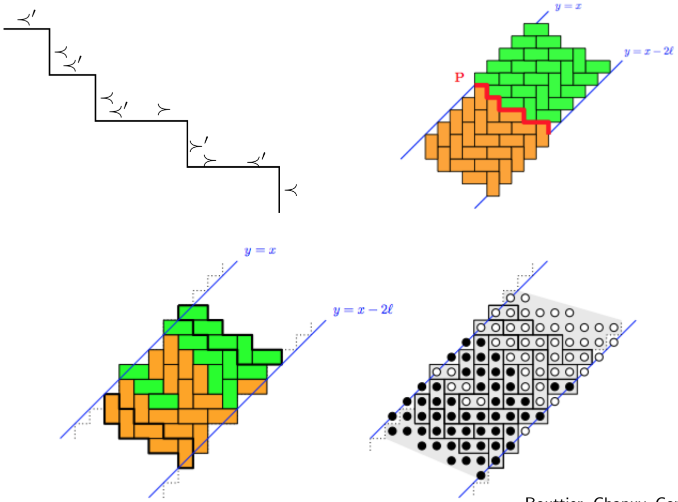
Pyramid partitions



$$w = (\underbrace{\ldots, \prec, \prec', \prec, \prec'}_{l}, \underbrace{\succ, \succ', \succ, \succ', \ldots}_{l})$$

Steep tilings

$w \in \{\prec, \succ, \prec', \succ'\}^{2l}$ $w_{2i} \in \{\prec, \succ\}$ and $w_{2i+1} \in \{\prec', \succ'\}$

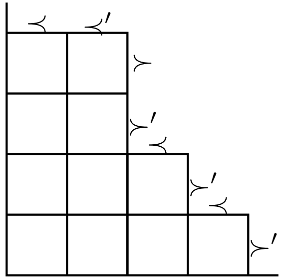


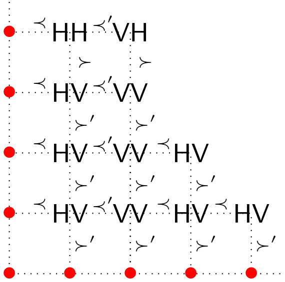
Bouttier, Chapuy, Corteel (2014)

Algorithm

Ex. $w = (\prec, \prec', \succ, \succ', \prec, \succ', \prec, \succ')$

- shape: path of horizontal ($w_i \in \{\prec, \prec'\}$) and vertical ($w_i \in \{\succ, \succ'\}$) segments
- : type: HH ($\prec, \succ$), HV ($\prec, \succ'$), VH ($\prec', \succ$), VV ($\prec', \succ'$)





Cauchy identity

$$\sum_{\nu} s_{\nu/\lambda}(x) s_{\nu/\mu}(y) = \frac{1}{1-xy} \sum_{\kappa} s_{\lambda/\kappa}(y) s_{\mu/\kappa}(x)$$

$$\begin{array}{ccc} \lambda & & \\ \curlyvee & G & \\ \kappa & \prec & \mu \end{array} \longrightarrow \begin{array}{ccc} \lambda & \prec & \nu \\ & & \curlyvee \\ & & \mu \end{array} \qquad \begin{tikzcd} \lambda \arrow[r] \arrow[d] \arrow[dr, phantom, "G"] & \nu \arrow[d] \\ \kappa \arrow[r] & \mu \end{tikzcd}$$

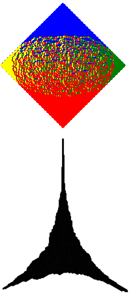
- sample $G \sim Geom(xy)$
- $\nu_i = \begin{cases} \max(\lambda_1, \mu_1) + G & \text{if } i = 1, \\ \max(\lambda_i, \mu_i) + \min(\lambda_{i-1}, \mu_{i-1}) - \kappa_{i-1} & \text{if } i > 1 \end{cases}$

Dual Cauchy identity

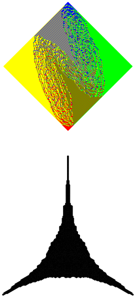
$$\sum_{\nu} s_{\nu/\lambda}(x)s_{\nu'/\mu'}(y) = (1+xy)\sum_{\kappa} s_{\lambda'/\kappa'}(y)s_{\mu/\kappa}(x)$$

$$\begin{array}{ccc} \lambda & & \\ \curlyvee & & \\ \kappa & \prec & \mu \end{array} \quad \longrightarrow \quad \begin{array}{ccc} \lambda & \prec & \nu \\ & & \curlyvee \\ & & \mu \end{array} \qquad \begin{tikzcd} \lambda \arrow[r] \arrow[d] \arrow[dr, phantom, "B"] & \nu \arrow[d] \\ \kappa \arrow[r] & \mu \end{tikzcd}$$

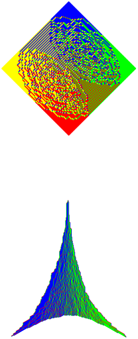
- sample $B \sim Bernoulli(\frac{xy}{1+xy})$
- for $i = 1 \ldots \max(\ell(\lambda), \ell(\mu)) + 1$
  if $\lambda_i \leq \mu_i < \lambda_{i-1}$ then $\nu_i = \max(\lambda_i, \mu_i) + B$
  else $\nu_i = \max(\lambda_i, \mu_i)$
  if $\mu_{i+1} < \lambda_i \leq \mu_i$ then $B = \min(\lambda_i, \mu_i) - \kappa_i$



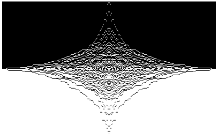
plane partition



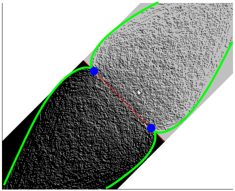
symmetric plane partition



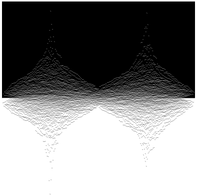
plane overpartition



symmetric pyramid partition



finite width pyramid partiton



steep tilings

Algorithm 1 Sampling from a DPP

Input: eigendecomposition $\{(\boldsymbol{v}_n, \lambda_n)\}_{n=1}^{N}$ of L
$J \leftarrow \emptyset$
for $n = 1, 2, \ldots, N$ **do**
  $J \leftarrow J \cup \{n\}$ with prob. $\frac{\lambda_n}{\lambda_n + 1}$
end for
$V \leftarrow \{\boldsymbol{v}_n\}_{n \in J}$
$Y \leftarrow \emptyset$
while $|V| > 0$ **do**
  Select i from $\mathcal{Y}$ with $\Pr(i) = \frac{1}{|V|} \sum_{\boldsymbol{v} \in V} (\boldsymbol{v}^\top \boldsymbol{e}_i)^2$
  $Y \leftarrow Y \cup i$
  $V \leftarrow V_\perp$, an orthonormal basis for the subspace of V orthogonal to $\boldsymbol{e}_i$
end while
Output: Y

Algorithm due to Hough et al. [2006]

Figure taken from Kulesza and Taskar (20012)