1. Exchangeable random variables & Polya urn scheme.

1.1. De Finetti theorem. infinite sequences of exchangeable random variables

Def

$\{X_1 \dots X_n, \dots\}$ is exchangeable if for $\forall$ finite permutation $\sigma \in S_n$ $n<\infty$, $\{X_i\}_{i=1}^n \sim \{X_{\sigma(i)}\}_{i=1}^n$ (in law)

Example: iid, or: mixture of iid is okay.

Theorem de Finetti. $\{X_i\}_{i=1}^\infty$ exchangeable sequence of r.v.-s, $X_i \in \{0,1\}$. Then $\exists$ random variable U, $\mathbb{P}(U\in[0,1])=1$ s.t. $\forall$ $1 \le k \le n$ $\mathbb{P}(X_1=X_2=\dots=X_k=1,\ X_{k+1}=\dots=X_n=0)=$

$$= \mathbb{E}\left(U^k(1-U)^{n-k}\right)$$

basically: (conditionally independent Bernoulli-s: random success prob. U) (holds in more general as well)

$S_n=\sum_{k=1}^n X_k$; $\mathbb{P}(S_n=k)=\mathbb{E}\left(\mathbb{P}(BIN(n,U)=k)\right)=\mathbb{E}\left(\binom{n}{k}\cdot U^k(1-U)^{n-k}\right)$

$\Downarrow$ we can compute the distribution of U:

$\lim_{n\to\infty} n\cdot\mathbb{P}(S_n=\lceil un\rceil)=f(u)$ (Strong LLN applied for conditional distribution given $U=u$ $\frac{S_n}{n}\xrightarrow{a.s.}u \rightsquigarrow \mathbb{P}\left(\frac{S_n}{n}\in(u,u+du)\right)=f(u)du$

Proof: Quite simple.

fix $m \ge n$, condition on S_m

Helly's thm: $\forall$ bounded sequence of r.v.s have a weakly convergent subsequence. (tight $\to$ Prohorov)

$\mathbb{P}(X_1=\dots=X_k=1,\ X_{k+1}=\dots=X_n=0)=$

$$=\sum_{j=k}^m \mathbb{P}(X_1=\dots=X_k=1, X_{k+1}=\dots=X_n=0 \mid S_m=j)\cdot\mathbb{P}(S_m=j)$$

by exchangeability each sequence with j db 1-s is equally likely $\frac{1}{\binom{m}{j}}$, and the initial config. 111…10000 $j-k$ és $m-n$ helyre

$$\sum_{j=k}^m \frac{\binom{m-n}{j-k}}{\binom{m}{j}}\mathbb{P}(S_m=j)=\sum_{j=k}^m \frac{j(j-1)\cdots(j-k+1)\cdot(m-j)\cdots(m-j-(n-k)+1)}{m(m-1)\cdots(m-n+1)}\cdot\mathbb{P}(S_m=j)$$

$m\to\infty$, $k<n$ fix $\longrightarrow \sum_{j=k}^\infty \left(\frac{j}{m}\right)^k\cdot\left(1-\frac{j}{m}\right)^{n-k}+o(1)\cdot\mathbb{P}(S_m=j)=$

$Y_m:=\frac{S_m}{m}$

$0\le\frac{S_m}{m}\le\frac{m}{m}=1$

$$=\sum_{j=k}^\infty (Y_m)^k\cdot(1-Y_m)^{n-k}\cdot\mathbb{P}\left(Y_m=\frac{j}{m}\right)=\mathbb{E}\left(Y_m^k\cdot(1-Y_m)^{n-k}\right)$$

$\boxed{\text{Infinite sequence}}$

$\{Y_m\}$ bounded $\Rightarrow$ $\exists$ weakly conv. subsequence, $\{Y_{m_\ell}\}_{\ell=1}^\infty \xrightarrow{d} U$

$$\lim_{m\to\infty}\mathbb{E}\left(Y_m^k(1-Y_m)^{n-k}\right)\overset{\text{uniformly bounded}}{=}\lim_{\ell\to\infty}\mathbb{E}\left((Y_{m_\ell})^k(1-Y_{m_\ell})^{n-k}\right)=\mathbb{E}\left(U^k(1-U)^{n-k}\right).$$

Polya - urn Schemes (Appl. of deFinetty

$\boxed{b_0 \quad r_0}$ blue red

$W_b : \mathbb{N} \to (0,\infty)$ weight function ; $\mathbb{P}(\text{Blue}) = \dfrac{W_b(B_n)}{W_b(B_n) + W_r(R_n)}$

$W_r : \mathbb{N} \to (0,\infty)$ $\boxed{B_n \; R_n}$

(typical: linear: $W_b(k) = a_b + k$, $W_r(k) = a_r + k$ $\boxed{\text{second ball of the same color}}$

same slope, possibly different intercept.

Thm. $\{(B_n, R_n)\}_{n=1}^{\infty}$ Polya urn scheme with linear W_b & W_r with the same slope. Then, $\dfrac{B_n}{B_n + R_n} \xrightarrow{a.s.} U$, $U \sim \text{Beta}(a, b)$

$a = b_0 + a_b$
$b = r_0 + a_r$

$$\mathbb{P}(B_n = B_0 + k) = \mathbb{E}\left(\mathbb{P}(BIN(n, U) = k)\right) = \mathbb{E}\left[\binom{n}{k} \cdot U^k (1-U)^{n-k}\right].$$

$$f_{a,b}(x) = \frac{x^{a-1} \cdot (1-x)^{b-1}}{\underbrace{\int_0^1 x^{a-1}(1-x)^{b-1}dx}_{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}}}$$

Proof. $X_n = \mathbb{1}(n\text{-th ball drawn blue})$

Show: X_n infinite, exchangeable

$$B_n = b_0 + \sum_{j=1}^n X_j. \quad R_n = r_0 + \sum_{j=1}^n (1 - X_j) = r_0 - b_0 + n - B_n$$

$\{x_t\}_{t=1}^n$ 0–1 sequence $\quad b_t = b_0 + \sum_{j=1}^t x_j$. $\quad r_t = r_0 - b_0 + t - b_t$

$$\mathbb{P}\left(\{X_t\}_{t=1}^n = \{x_t\}\right) = \prod_{t=1}^n \frac{W_b(b_t)^{x_t} \cdot W_r(r_t)^{1-x_t}}{W_b(b_t) + W_r(r_t)} = *$$

nevező: $\prod_{t=1}^n \left(W_b(b_t) + W_r(r_t)\right) = \prod_{t=0}^{n-1} (b_0 + r_0 + a_b + a_r + t)$ $\quad k = \sum_1^n x_t$

nababló $\prod_{t=1}^n W_b(b_t)^{x_t} = \prod_{m=0}^{k-1}(b_0 + a_b + m)$; $\prod_{t=1}^n W_r(r_t)^{1-x_t} = \prod_{j=0}^{n-k-1}(r_0 + a_r + j)$

$$* = \frac{\prod_{m=0}^{k-1}(\tilde{b} + m) \cdot \prod_{j=0}^{n-k-1}(\tilde{r} + j)}{\prod_{t=0}^{n-1}(\tilde{b} + \tilde{r} + t)}$$

nem függ a sorrendtől, csak az 1-esek totális számától
⇒ infinite exchangeable sequence.
⇒ mixture of Bernoullis, random success probability U.
Distribution depends on b_0, r_0, a_b, a_r
$\tilde{b} = b_0 + a_b \quad \tilde{r} = r_0 + a_r$

Density calculation

$\lim_{n\to\infty} n\mathbb{P}(S_n = \lceil un \rceil) = f(u)$. Fix $0 \le k \le n$, $\binom{n}{k}$ sequence with k 1-s, each has probab. ⊛

$$\mathbb{P}(S_n = k) = \binom{n}{k} \cdot ⊛ = \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \cdot \frac{\Gamma(k+\tilde{b})}{\Gamma(\tilde{b})} \cdot \frac{\Gamma(n-k+\tilde{r})}{\Gamma(\tilde{r})} \cdot \frac{\Gamma(\tilde{b}+\tilde{r})}{\Gamma(\tilde{b}+\tilde{r}+n)} =$$

$$= \frac{\Gamma(\tilde{b}+\tilde{r})}{\Gamma(\tilde{b}) \cdot \Gamma(\tilde{r})} \cdot \frac{\Gamma(k+\tilde{b})}{\Gamma(k+1)} \cdot \frac{\Gamma(n-k+\tilde{r})}{\Gamma(n-k+1)} \cdot \frac{\Gamma(n+1)}{\Gamma(n+\tilde{b}+\tilde{r})} =$$

$$\frac{\Gamma(\tilde{b}+\tilde{r})}{\Gamma(\tilde{b}) \cdot \Gamma(\tilde{r})} \cdot k^{\tilde{b}-1} \cdot (n-k)^{\tilde{r}-1} \cdot \frac{1}{n^{\tilde{b}+\tilde{r}-1}} =$$

$k = \lceil un \rceil$

$$\frac{\Gamma(\tilde{b}+\tilde{r})}{\Gamma(\tilde{b})\Gamma(\tilde{r})} \cdot u^{\tilde{b}-1} \cdot (1-u)^{\tilde{r}-1} \cdot \frac{1}{n}$$

✓ beta( ) kell.

Applications to scale free trees

Preferential attachment model $PA_t(1,\sigma)$

Recall

$t=2$ We do not allow loops, so

$$\mathbb{P}(v_{t+1} \to v_i) = \frac{D_i(t)+\sigma}{t(2+\sigma)+(d_1-2)+(d_2-2)}$$

Decomposition into two trees:
$T_1(t)$: vertices closer to 1 than to 2 : $T_2(t)$: other.

$S_i(t) = |T_i(t)|$ $\quad S_1(t)+S_2(t) = t$

Thm. 1: $\displaystyle \frac{S_1(t)}{S_1(t)+S_2(t)} \xrightarrow{a.s.} U \qquad U \sim \text{Beta} \quad a = \frac{d_1+\sigma}{2+\sigma}$

$\displaystyle b = \frac{d_2+\sigma}{2+\sigma}$

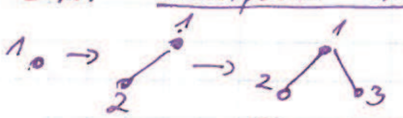
two scale free trees, each of them positive proportion of vertices, proportion converges to Beta distribution.

$\mathbb{P}(S_1(t)=k) =$
$= \mathbb{E}\left(\mathrm{BIN}(t-1,U)=k-1\right)$

Proof: $\displaystyle S_1(t)=s_1(t);\ \mathbb{P}(v_{t+1}\to T_1(t)) = \frac{(2s_1(t)+d_1-2)+\sigma\cdot s_1(t)}{(2s_1(t)+d_1-2)+\sigma s_1(t)+(2s_2(t)+d_2-2)+\sigma s_2(t)}$

$$= \frac{s_1(t)+\frac{d_1-2}{2+\sigma}}{s_1(t)+s_2(t)+\frac{d_1+d_2-4}{2+\sigma}} \rightsquigarrow r_0=b_0:=1 \quad a_b = \frac{d_1-2}{2+\sigma} \quad a_r = \frac{d_2-2}{2+\sigma} \qquad \square$$

Ex. Uniform recursive tree



attach v_{t+1} uniformly to $v_1 \dots v_t$.

Prove that $\displaystyle \frac{S_1(t)}{S_1(t)+S_2(t)} \xrightarrow{a.s.} U \qquad U \sim U[0,1]$.

$\displaystyle \mathbb{P}(S_1(t)=k) = \frac{1}{t}$.

Thm 2. Relative degrees in Scale-free trees.

$PA_t(1,\sigma)$ $\quad D_i(t)$ denotes degree of vertex i at time t.

For any fixed k, $\displaystyle \frac{D_k(t)}{D_1(t)+\dots+D_k(t)} \xrightarrow{a.s.} B_k$

$\displaystyle B_k \sim \text{Beta} \quad a=1+\sigma \quad b=(k-1)(2+\sigma) \qquad \left(B_k = \frac{\xi_k}{\xi_1+\dots+\xi_k}\right)$

Proof. Stopping times: $\{T_k(n)\}_{n=2k-1}^{\infty}$ $\quad T_k(2k-1)=k-1$
fix k.

$T_k(n) = \inf\{t : D_1(t)+\dots+D_k(t)=n\}$

$T_k(n) < \infty \ \forall n \quad (D_i(t)\to\infty\ \forall i)$
but $\lim_{n\to\infty} T_k(n) = \infty$

$$\lim_{t\to\infty} \frac{D_k(t)}{D_1(t)+\dots D_k(t)} = \lim_{n\to\infty} \frac{D_k(T_k(n))}{D_1(T_k(n))+\dots D_k(T_k(n))} = \lim_{n\to\infty}\frac{D_k(T_k(n))}{n}.$$

subsequence

Q: What is the distribution of $D_k(T_k(n))$ when total degree of first k vertices is n?

$\{ D_k(T_k(n)), (D_1 + \dots D_{k-1}(T_k(n))) \}_{n=2k-1}^{\infty}$ Polya-urn scheme

$D_k(T_k(2k-1)) = 1$, $D_1 + \dots D_{k-1}(T_k(2k-1)) = 2k-2$

at each $T_k(.)$, a new edge is attached to D_k or to $D_1 + \dots D_{k-1}$

$\mathbb{P}(V_{T_k(n)} \to V_k) = \dfrac{D_k(T_k(n)) + \sigma}{n + k\sigma}$ $\rightsquigarrow$ $a_b = \sigma$ $a_r = (k-1)\cdot\sigma$, $b_0 = 1$ $r_0 = 2k-2$.

$\Rightarrow$ Claim follows. ✓

Connectivity of $PA_t(1,\sigma)$, $PA_t(m,\sigma)$ $m \geq 1$

$m=1$ special $\rightsquigarrow$ if you draw a loop, that part stays disconnected forever.

$N_t = l_1 + l_2 + \dots l_t$ $\qquad \mathbb{P}(l_j = 1) = \dfrac{1+\sigma}{(2+\sigma)(j-1) + 1+\sigma}$ $\qquad \sum \mathbb{P}(A_j) = \infty$ + indep.

$\Rightarrow$ Borel Cantelli lemma implies that infinitely many l_j-s are 1 $\Rightarrow N_t \to \infty$

Moreover, not hard to see $\dfrac{N_t}{\log t} \xrightarrow{\mathbb{P}} \dfrac{1+\sigma}{2+\sigma} < 1$

$\Rightarrow$ with high prob, $\exists$ largest component, size $\geq \dfrac{t}{\log t}$

($N_t \approx \underbrace{\frac{1+\sigma}{2+\sigma}}_{<1} \log t$ db komponent van t csúcson $\rightarrow$

Ex.: N_t in $PA_t(1,\sigma)$ satisfies CLT: $\mathbb{E}N_t = \dfrac{1+\sigma}{2+\sigma}\log t \cdot (1+o(1))$

$\operatorname{Var} N_t = \dfrac{1+\sigma}{2+\sigma} \log t \, (1+o(1))$

$m \geq 1$.

Thm. Connectivity for $m \geq 2$. Whp, $PA_t(m,\sigma)$ connected.

Proof. new component $l_t = N_t - N_{t-1} \Leftrightarrow$ ~~of~~ all m edges are attached to the same vertex.

$\mathbb{P}(l_t = 1) = \prod_{j=1}^{m} \dfrac{2j-1+\sigma}{(2m+\sigma)t + (2j-1+\sigma)} \approx \dfrac{c(m)}{t^m + d(m)}$

$m > 2$

$\Rightarrow \sum_{t=2}^{\infty} \mathbb{P}(l_t = 1) < \infty \Rightarrow$ a.s. $\{l_t = 1\}$ happens finitely many times.

$\Rightarrow N_t = 1 + \sum l_t < \infty$ a.s.

$\Rightarrow$ a.s. we have finite number of components.

Extension. $\mathbb{P}(N_t = 1) \xrightarrow[t\to\infty]{} 1$

Proof. fix K large enough s.t. $\sum_{t=K}^{\infty} l_t = 0$ (happens w.p. $\to 1$, condition on this)

$\to$ no new component after K. $\{v_1 \dots v_K\}$.

Proof. 1. Whp. # vertices connected to $i \in \{1 \dots K\}$ is large ($C_i(t)$)

2. implies in fact all vertices connected to i.

Polya-urn with $b_0 = 1$, $r_0 = i$ $\quad (B_n, R_n)$: weight: proportional, blue $\rightsquigarrow$ add blue, red $\rightsquigarrow$ add red

Claim

$C_i(t) \geq B_{t-i}$. Proof by induction.

time i, $C_i(i) \geq 1 \quad B_{i-i} = B_0 = b_0 = 1$ ✓

$C_i(t) \geq B_{t-i} \rightsquigarrow t+1$

total weight of vertices in $C_i(t)$ $\geq B_{t-i}(2m+\delta)$

at time $t+1$ we draw a blue at $t-i+1$ if first edge of $v_{t+1} \longrightarrow v_y : v_y \in C_i(t)$

total weight $= t(2m+\delta) + (1+\delta) \leq$
$\leq (t+1)(2m+\delta) =$
$= (\underbrace{B_{t-i} + R_{t-i}}_{t+1})(2m+\delta)$

$C_i(t+1) \geq \hookrightarrow \mathbb{P} \geq \frac{w(C_i(t))}{\text{total } w.} \geq$

drawing a blue $\geq \frac{B_{t-i}}{B_{t-i} + R_{t-i}}$ ✓

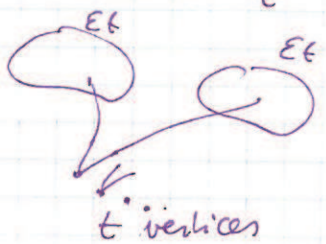
$\Rightarrow C_i(t+1) \geq B_{t-i+1}$ ✓ $\quad \frac{B_t}{t} \xrightarrow{a.s} U \quad U \sim \beta eta(a=1, b=i)$

$\mathbb{P}(U=0) = 0$

$\Rightarrow$ w.h.p. $C_i(t) \geq \varepsilon(t-i)$ $\forall t$ large, ε suff. small. $\quad \frac{B_{t-i}}{t-i} \geq \varepsilon$

$\mathbb{P}\left(\underline{\lim} \frac{C_i(t)}{t} \geq \varepsilon \ \forall i = 1 \dots K\right) = 1 - o(1)$ as $\varepsilon \to 0$

2nd part. $\frac{C_i(t)}{t} \geq \varepsilon$, $\frac{C_j(t)}{t} \geq \varepsilon$ $i \neq j \Rightarrow i \sim j$ connected in PA_{2t} at $2t$



$\ell = \{t+1, \dots 2t\}$

If $\exists \ell$: first edge into $C(i)$, second into $C(j)$, $\Big\}$ $i \longleftrightarrow j$ connected via ℓ.

[ℓ is a t-connector for i,j]

$\mathbb{P}(\ell \text{ is a } t\text{-connector}) \geq \frac{C_i(t)(2m+\delta)}{\ell \cdot (2m+\delta) + 1 + \delta} \cdot \frac{C_j(t) \cdot (2m+\delta)}{\ell(2m+\delta) + 2 + \delta} \geq \frac{C_i(t) C_j(t)}{(2t+1)^2} \geq \frac{\varepsilon^2}{16}$

independently of other $t+1 \dots \ell-1$ vertices.

$\Rightarrow \mathbb{P}(\text{no-}t\text{-connector}) \leq \left(1 - \frac{\varepsilon^2}{16}\right)^t \leq e^{-\frac{\varepsilon^2}{16} t} \to 0$ exponentially $t \to \infty$.