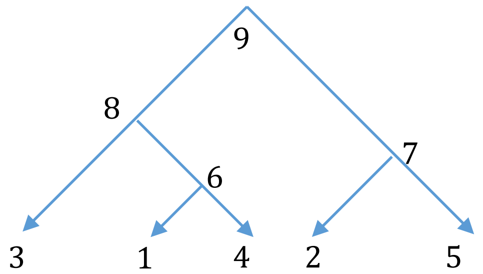
۱ شبیه‌سازی coalescent

برای شبیه‌سازی درخت coalescent، گره‌های داخلی درخت را نیز نام‌گذاری می‌کنیم و الگوریتم را برای یک نمونه‌ی nتایی با $V_0 = \{1, \dots, n\}$ و با $T_n = 0$ شروع می‌گردد. شکل زیر نمونه‌ای از شبیه‌سازی درخت از پایین را نمایش می‌دهد.



$$V_3 \to \{7,8\}$$
$$V_2 \to \{3,6,7\}$$
$$V_1 \to \{2,3,5,6\}$$
$$V_0 \to \{1,2,3,4,5\}$$

۱-۱ الگوریتم ساخت درخت از پایین

قرار بده $V_0 = \{1, \dots, n\}$ و $T_n = 0$ و برای $k = 0,1, \dots, n-2$ انجام بده:

- دو عدد تصادفی i_k و j_k را از V_k انتخاب کن
- قرار بده $V_{k+1} = (V_k - \{i_k, j_k\}) \cup \{n+k+1\}$
- در درخت دو گره‌ی i_k و j_k را به $n+k+1$ وصل کن
- t_{n-k} را یک نمایی مستقل با متوسط $\frac{1}{\binom{n-k}{2}}$ قرار بده (توزیع نمایی با پارامتر $\lambda = \binom{n-k}{2}$)
- قرار بده $T_{n-k-1} = T_{n-k} + t_{n-k}$

نکته: توزیع نمایی را به سادگی می‌توان با توزیع یکنواخت و به صورت

$$t_{n-1} = \frac{1}{\binom{n-k}{2}} \ln(\frac{1}{U_k})$$

که U_kها مستقل می‌باشند شبیه‌سازی نمود.

برای آن‌که بتوانیم جهش را شبیه‌سازی نماییم، می‌توانیم تابع والد (ancestor) را به صورت

$$anc[i_k] = n+k+1 \, , anc[j_k] = n+k+1$$

استفاده کنیم. بدین ترتیب تعداد جهش‌های روی یک یال که دارای توزیع پواسون با متوسط $\frac{\theta}{2}(T_{anc[i]} - T_i)$ می‌باشد را شبیه‌سازی نماییم ($\theta = 4N\mu$).

نگارنده: محمد امیرخانی
ایمیل: m_amirkhani@ce.sharif.edu

۲ ساخت درخت coalescent از بالا (genealogical)

به طور کلی می‌توانیم درخت را به صورت دیگری نیز نمایش دهیم. در مرحله‌ی T_1، کل افراد در یک مجموعه‌ی واحد $\{1, \dots, n\}$ قرار می‌گیرند و در مرحله‌ی T_2 عناصر به دو مجموعه تفکیک گردیده و T_3 به سه مجموعه و الی آخر.

$$
\begin{aligned}
&T_1 \to \{1,2,3,4,5\} \\
&T_2 \to \{1,3,4\}\{2,5\} \\
&T_3 \to \{1,4\}\{2,5\}\{3\} \\
&T_4 \to \{1,4\}\{2\}\{3\}\{5\} \\
&T_5 \to \{1\}\{2\}\{3\}\{4\}\{5\}
\end{aligned}
$$

اگر ε_n را مجموعه‌ی تمام تقسیم‌های ممکن برای $\{1, \dots, n\}$ قرار دهیم، به ازای هر $\xi \in \varepsilon_n$، $|\xi|$ را تعداد مجموعه‌هایی که ξ را تشکیل می‌دهند قرار می‌دهیم. هم‌چنین تفکیک‌های زمان T_i را که در آن برای اولین بار i شاخه[۱] بوجود آمده است را به صورت ξ_i^n نمایش می‌دهیم.

۲-۱ قضیه (Kingman 1982a):

اگر ξ یک تفکیک مجموعه‌ی $\{1, \dots, n\}$ باشد که برای آن $|\xi| = i$ است، آنگاه

$$P(\xi_i^n = \xi) = C_{n,i} w(\xi)$$

که در آن

$$w(\xi) = \lambda_1! \dots \lambda_i!$$

است و λ_i ها تعداد عناصر مجموعه‌های تفکیک شده می‌باشند. هم‌چنین ثابت $C_{n,i}$ به صورت

$$C_{n,i} = \frac{i!}{n!} \times \frac{(n-i)!\ (i-1)!}{(n-1)!}$$

انتخاب می‌شود که مجموع احتمالات برابر یک گردد.

اثبات: از استقرا برای اثبات قضیه Kingman استفاده می‌کنیم.

پایه‌ی استقراء: در حالت پایه $|\xi| = i = n$ است و تفکیک $\{1, \dots, n\}$ به n زیرمجموعه تنها به شکل $\xi = \{\{1\}, \dots, \{n\}\}$ ممکن است که به احتمال ۱ اتفاق میفتد. از طرفی برای این حالت $C_{n,i}$ و $w(\xi)$ هردو برابر یک است و لذا رابطه در حالت پایه‌ی استقرا برقرار می‌باشد.

فرض استقراء:

---

[۱] lineage

نکته: وزن های $w(\xi)$ به گونه‌ای تعریف شده است که به تفکیک‌های غیرمساوی وزن بیشتری می‌دهد. به عنوان مثال برای یک مجموعه‌ی ۹ عضوی در حالتی که $i = 3$ باشد، تفکیک‌های زیر با وزن‌های داده شده بدست می‌آید.

$$\begin{cases} 3-3-3 \rightarrow 216 \\ 4-3-2 \rightarrow 288 \\ 5-2-2 \rightarrow 480 \\ 4-4-1-576 \\ 5-3-1 \rightarrow 720 \\ 6-2-1 \rightarrow 1440 \\ 7-1-1 \rightarrow 5040 \end{cases}$$

۲-۲ قضیه:

اگر π یک جایگشت تصادفی از $\{1, \dots, n\}$ بوده و $\lambda_j = |\xi_{i,\pi(j)}^{n}|$ باشد که در آن $\pi(j)$ عنصر jام از جایگشت تصادفی π باشد، آنگاه $(\lambda_1, \dots, \lambda_i)$ به طور یکنواخت بر روی بردارهای مثبتی که دارای مجموع n می‌باشند توزیع می‌گردد.

اثبات: اگر مجموعه‌های داخل ξ را به صورت تصادفی ردیف نماییم، هرکدام از ردیف‌ها دارای احتمال

$$\frac{C_{n,i} w(\xi)}{i!}$$

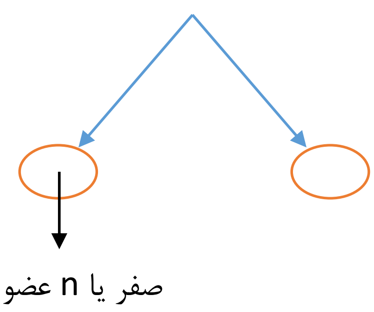
می‌باشند. از طرفی تعداد حالت‌هایی که می‌توانند بردار $(\lambda_1, \dots, \lambda_i)$ را ایجاد نمایند برابر با

$$\frac{n!}{\lambda_1! \dots \lambda_i!}$$

است. با ضرب کردن این دو کمیت به احتمال $\frac{1}{\binom{n-1}{i-1}}$ برای یک ردیف خاص با تعداد عناصر $(\lambda_1, \dots, \lambda_i)$ می‌رسیم. چون کمیت تنها به i و n ربط داشته و مستقل از λ_j می‌باشد، پس توزیع یکنواخت است. تعداد جواب‌های مثبت معادله‌ی $\sum_{j=1}^{i} \lambda_j = n$ برابر $\binom{n-1}{i-1}$ خواهد بود.

۲-۳ قضیه

احتمال آنکه جد مشترک یک جمعیت با جد مشترک یک نمونه‌ی nتایی آن برابر باشد با میل جمعیت به سمت بی‌نهایت معادل $\frac{n-1}{n+1}$ خواهد بود.**اثبات:** اگر جدها بخواهند متفاوت باشند، بایستی از مرحله‌ی اول تفکیک‌پذیری جمعیت، هر n عنصر نمونه در یک شاخه قرار گیرند.



اگر $n = x_1 + x_2$ باشد، دو جواب $x_1 = 0, x_2 = n$ و $x_1 = n, x_2 = 0$ مطلوب هستند که احتمال آن برابر $\frac{2}{n+1}$ است. اگر x را نسبت تعداد نمونه‌ها در مجموعه‌ی اول بدانیم، دارای توزیع یکنواخت بوده و

$$\int_0^1 x^n + (1-x)^n \, dx = \frac{2}{n+1}$$

و احتمال یکسان بودن برابر است با $1 - \frac{2}{n+1}$ که نتیجه‌ی دلخواه می‌باشد.

۲-۴ قضیه (ساخت درخت coalescent از بالا):

برای آن‌که ξ_i^n را از روی $\{A_1, \dots, A_{i-1}\} = \xi_{i-1}^n$ بسازیم، یکی از مجموعه‌ها را با احتمال انتخاب

$$P[A_j] = \frac{\lambda_j - 1}{n - i - 1}$$

که $\lambda_i = |A_j|$ به طور تصادفی انتخاب می‌کنیم و A_j را به دو مجموعه با اندازه‌های k و $\lambda_j - k$ که در آن k توزیع یکنواخت بر روی $1,2, \dots, \lambda_j - 1$ دارت تفکیک می‌نماییم.

اثبات: کافیست $P[\xi_i^n = \xi | \xi_{i-1}^n = \eta]$ را محاسبه نماییم. خواهیم داشت

$$P[\xi_i^n = \xi | \xi_{i-1}^n = \eta] = \frac{P[\xi_i^n = \eta | \xi_{i-1}^n = \xi] P[\xi_i^n = \xi]}{P[\xi_{i-1}^n = \eta]}$$

که تمام احتمالات را قبلا محاسبه نموده‌ایم. لذا با جاگذاری داریم:

$$P[\xi_i^n = \xi | \xi_{i-1}^n = \eta] = \frac{1}{n-i+1} \times \frac{k! \left(\lambda_j - k\right)!}{\lambda_j!} \times 2$$
$$= \frac{\lambda_j - 1}{n-i+1} \times \frac{1}{\lambda_j - 1} \times \frac{2}{\binom{\lambda_j}{k}}$$

که سه جمله به ترتیب (از چپ)، احتمال انتخاب A_j، احتمال تفکیک k و تخصیص عناصر به مجموعه‌ی kتایی است.

۳ مدل wright-fisher دو Allele و Fixation:

صفحات ۲۲ تا نیمه‌ی ۲۵ توسط استاد تدریس نشد!

۳-۱ Hetrozygosity

احتمال آن که یک جفت انتخاب شده در مرحله‌ی nام با یکدیگر متفاوت باشند.

$$H_n^0 = \frac{X_n (2N - X_n)}{\binom{2N}{2}}$$

۳-۲ قضیه:

قرار دهید $h(n) = E[H_n^0]$ که متوسط Hetrozygosity در مرحله‌ی nام می‌باشد. در مدل Wright-Fisher

$$h(n) = \left(1 - \frac{1}{2N}\right)^n h(0)$$

می‌باشد.

اثبات۱: با توجه به آن‌که یک توزیع دوجمله‌ای دارای خواص

$$E[x] = np\, ; E[X^2] = n^2p^2 + np(1-p)$$

است، می‌توان به سادگی فهمید که

$$\begin{aligned} E[H_n^0|X_{n-1}] &= \frac{2NE[X_n|X_{n-1}] - E[X_n^2|X_{n-1}]}{\binom{2N}{2}} \\ &= \frac{2N\,X_{n-1} - X_{n-1}^2 - X_{n-1}\left(1 - \frac{X_{n-1}}{2N}\right)}{\binom{2N}{2}} \\ &= \left(1 - \frac{1}{2N}\right)\frac{X_{n-1}(2N - X_{n-1})}{\binom{2N}{2}} \\ &= \left(1 - \frac{1}{2N}\right)H_{n-1}^0 \end{aligned}$$

و امید گرفتن از طرفین به جواب مطلوب می‌رسیم.

اثبات۲: اگر دو موجود را در مرحله‌ی nام نگاه کنیم، احتمال آن‌که این دو متفاوت باشند به آن باز می‌گردد که والدین آن‌ها متفاوت باشند و این احتمال برابر $1 - \frac{1}{2N}$ است. این عمل بایستی تا مرحله‌ی آغازین پیش رفته $\left(1 - \frac{1}{2N}\right)$ و در آن‌جا احتمال متفاوت بودن برابر با $h(0)$ است.