



# ژنومیک محاسباتی

مطهری و فروغمند  
پاییز ۱۴۰۰

## درخت Coalescent

جلسه دوازدهم

نگارنده: علیرضا صاحبی

### ۱. مروری بر مباحث گذشته

در جلسه گذشته که جلسه اول تدریس قسمت دوم درس توسط دکتر مطهری بوده است، مفهوم ژنتیک جمعیت به اختصار ذکر شد و درباره عوامل مختلف تفاوت های ژنوم موجودات صحبت شد. همچنین قانون Hardy-Weinberg معرفی شد که دارای شرایطی از جمله جمعیت Diploid و اندازه بینهایت است. سپس درباره مدل دیگری با نام Wright-Fisher صحبت شد که شرایط دیگری از جمله جمعیت Haploid و محدود و ثابت را در نظر میگیرد. در این جلسه، به ادامه مدل Wright-Fisher و معرفی درخت Coalescent خواهیم پرداخت.

### ۲. مدل Wright-Fisher

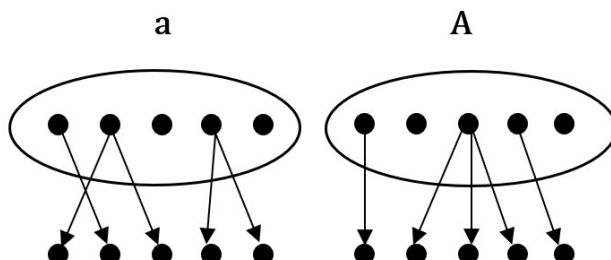
در جلسه گذشته مطرح شد که اگر جمعیت جامعه مورد بررسی محدود باشد، مدلی که میتوان با آن جمعیت را بررسی کرد، Wright-Fisher نام دارد. در این مدل انتخاب طبیعی<sup>۱</sup> وجود ندارد و تمام جهش ها، خنثی<sup>۲</sup> هستند. همچنین جمعیت ثابت بوده و اگر جمعیت را Haploid فرض کنیم، هرکدام از اعضای موجود در جمعیت فعلی، یکی از اعضای نسل قبل را به عنوان والد خود انتخاب میکند؛ یعنی اگر فرض کنیم جمعیت جامعه مورد بررسی  $2N$  است و یک عضو مشخص از نسل فعلی را در نظر بگیریم، هر کدام از اعضای نسل قبل به احتمال  $\frac{1}{2N}$  میتوانند به عنوان والد آن عضو مشخص باشد.

---

Selection<sup>۱</sup>  
Neutral<sup>۲</sup>

## ۱.۲ مدل سازی

فرض کنیم یک جمعیت Haploid با اندازه  $2N$  داریم که شرایط مدل Wright-Fisher در آن صدق میکند. همچنین برای یک ژن خاص در ژنوم اعضای جامعه مورد بررسی، دو نوع Allele متفاوت A و a وجود داشته باشد و در ابتدایی ترین نسل این جمعیت، تعداد اعضای که نوع A را دارند برابر با  $x$  باشد (شکل ۱).



شکل ۱: جامعه نمونه  $10$  عضوی دارای دو نوع Allele متفاوت A و a. هر عضو در نسل بعد، یک والد از نسل قبل دارد که نوع Allele خود را از آن به ارث میبرد.

متغیر  $X_i$ ، فرکانس Allele نوع A در نسل  $i + 1$  است. بنابراین فرکانس Allele های از نوع A در نسل اول برابر  $X_0 = \frac{x}{2N}$  است. احتمال اینکه  $k$  تا از عناصر نسل دوم دارای Allele نوع A باشند برابر است با:

$$P \left[ X_1 = \frac{k}{2N} \mid X_0 = \frac{x}{2N} \right] = \binom{2N}{k} \left( \frac{x}{2N} \right)^k \left( 1 - \frac{x}{2N} \right)^{2N-k}$$

احتمال اینکه  $m$  تا از عناصر نسل سوم دارای Allele نوع A باشند برابر است با:

$$P \left[ X_2 = \frac{m}{2N} \mid X_1 = \frac{k}{2N}, X_0 = \frac{x}{2N} \right] = P \left[ X_2 = \frac{m}{2N} \mid X_1 = \frac{k}{2N} \right] = \binom{2N}{m} \left( \frac{k}{2N} \right)^m \left( 1 - \frac{k}{2N} \right)^{2N-m}$$

در این احتمال، از تعداد افرادی که در نسل اول دارای Allele نوع A بودند، مستقل می‌شویم. به عبارت دیگر، دنباله  $X_i$  ها تشکیل یک زنجیره مارکوف میدهند که برای تعیین هر عضو از دنباله، تنها نیاز به عضو قبلی داریم؛ به عنوان مثال برای  $X_3$  نیز میتوان نوشت:

$$P \left[ X_3 \mid X_2, X_1, X_0 \right] = P \left[ X_3 \mid X_2 \right]$$

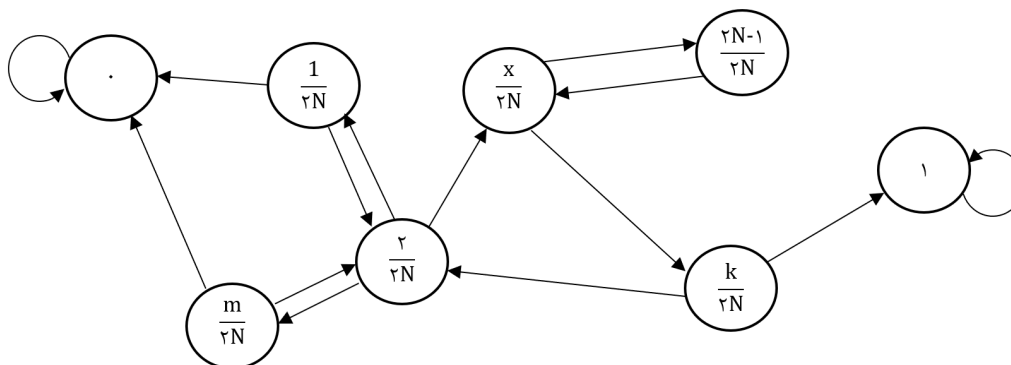
## ۲.۲ تثبیت<sup>۳</sup> و رانش<sup>۴</sup>

میدانیم  $X_i$  از جنس فرکانس است، پس مقدار هر  $X_i$  کسری است برابر با  $\frac{x}{2N}$  که  $x$  میتواند از  $0$  تا  $2N$  تغییر کند. بنابراین میتوان وضعیت های متفاوتی را برای هر  $X_i$  تعریف کرد که هرکدام از این وضعیت ها نشان دهنده یک فرکانس است که میتوان از هرکدام با یک احتمال به دیگری جابجا شد. شکل ۲ نمونه ای از این وضعیت ها و جابجایی بین آنها را نشان میدهد.

هنگامی که در نسل های مختلف حرکت میکنیم، مقدار فرکانس  $X_i$  بین این وضعیت های متفاوت جابجا میشود. نکته مهم این است که در این جابجایی ها، هرگاه وضعیت یک  $X_i$  به وضعیت صفر (فرکانس  $\frac{0}{2N}$ ) یا وضعیت یک (فرکانس  $\frac{2N}{2N}$ ) تغییر کند، دیگر تا همیشه وضعیت آن  $X_i$  در همان حال باقی خواهد ماند و تغییر نمیکند. از آنجا که برای ورود به هرکدام از این دو وضعیت یک احتمالی وجود دارد، بعد از گذشت چندین نسل، در نهایت وضعیت  $X_i$  به یکی از این دو تغییر میکند. بنابراین سرآخر یکی از Allele ها با ورود به وضعیت یک، تثبیت میشود (یعنی کل جامعه را فرا میگیرد) و Allele های دیگر، رانش پیدا خواهند کرد (یعنی از جامعه حذف میشود).

قضیه ۱. در یک جامعه که شرایط مدل Wright-Fisher در آن صدق میکند، اگر فرکانس Allele های نوع A در آن  $\frac{x}{2N}$  باشد، آنگاه احتمال اینکه تثبیت روی Allele نوع A اتفاق بیفتد، برابر است با:

$$P [Fix. is A] = \frac{x}{2N}$$



شکل ۲: یک نمونه از تغییر وضعیت‌های مختلفی که برای  $X_i$  می‌تواند اتفاق بیفتد. هنگام ورود به وضعیت با مقدار صفر و یا یک، تا همیشه در آن وضعیت باقی خواهد ماند.

اثبات. میدانیم که:

$$\mathbb{P}\left[X_1 = \frac{k}{2N} \mid X_0 = \frac{x}{2N}\right] = \binom{2N}{k} \left(\frac{x}{2N}\right)^k \left(1 - \frac{x}{2N}\right)^{2N-k}$$

امید ریاضی متغیر تصادفی  $X_1$  به شرط  $X_0$  برابر است با:

$$\mathbb{E}\left[X_1 \mid X_0 = \frac{x}{2N}\right] = \frac{x}{2N}$$

(به توالی از متغیرهای تصادفی  $V_0, V_1, V_2, \dots$  که  $\mathbb{E}[V_n \mid V_{n-1}, V_{n-2}, \dots, V_0] = \mathbb{E}[V_n \mid V_{n-1}]$  یعنی مقدار امید تنها به یک مرحله قبل وابسته است و از بقیه مستقل، Martingale گفته میشود)

از آنجا که دنباله  $X_i$  ها یک Martingale را تشکیل میدهند، میتوان نوشت:

$$\mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}\right] = \frac{x}{2N} \quad (1)$$

حال فرض کنیم  $n$  شماره یک نسل باشد. تثبیت یا قبل از  $n$  اتفاق افتاده، یا بعد از آن. بنابراین امید ریاضی رابطه (۱) را میتوان برحسب جمع دو حالت مختلفی که تثبیت قبل یا بعد از  $n$  اتفاق افتاده باشد، نوشت:

$$\begin{aligned} \mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}\right] &= \mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}, \text{Fix. before } n\right] \mathbb{P}[\text{Fix. before } n] \\ &+ \mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}, \text{Fix. after } n\right] \mathbb{P}[\text{Fix. after } n] \end{aligned}$$

این رابطه به ازای هر  $n$  برقرار است. حال اگر مقدار  $n$  را خیلی زیاد کنیم، احتمال اینکه تثبیت بعد از  $n$  اتفاق افتاده باشد صفر و احتمال اینکه تثبیت قبل  $n$  باشد برابر ۱ میشود. بنابراین خواهیم داشت:

$$\mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}\right] = \mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}, \text{Fix. before } n\right]$$

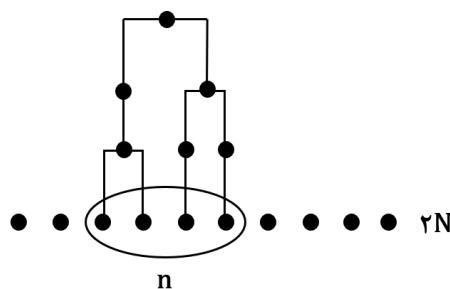
از طرفی با توجه به تعریف  $X_n$  میدانیم که اگر تثبیت روی  $a$  اتفاق بیفتد، مقدار  $X_n$  برابر خواهد بود با صفر و اگر روی  $A$  اتفاق بیفتد، برابر خواهد بود با ۱. بنابراین میتوان گفت امید شرطی مقدار  $X_n$  برابر خواهد بود با احتمال تثبیت روی  $A$ ، یعنی:

$$\mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}\right] = \mathbb{E}\left[X_n \mid X_0 = \frac{x}{2N}, \text{Fix. before } n\right] = \mathbb{P}[\text{Fix. is } A]$$

با توجه به رابطه (۱) قضیه اثبات میشود.

### ۳ درخت Kingsman Coalescent

فرض کنیم جامعه‌ای مانند شکل ۳ را به اندازه  $2N$  درحال بررسی هستیم و تعداد  $n$  نمونه (که  $n$  عدد کوچکی است) از جامعه انتخاب میکنیم. اگر شروع به حرکت به سمت نسل‌های قبلی کنیم، میتوان نسلی را پیدا کرد که دو عضو از این نمونه جد مشترک پیدا میکنند. به همین صورت اگر به سمت نسل‌های ابتدایی حرکت کنیم، میبینیم که دو به دوی اعضای نمونه با یکدیگر جد مشترک پیدا میکنند تا به نسلی با جد مشترک همه  $n$  نمونه‌ای که انتخاب کردیم برسیم. به درختی که جد‌های مشترک این  $n$  نمونه را به صورت مرحله به مرحله نشان میدهد، درخت Coalescent گفته میشود. در این قسمت میخواهیم راجع به ساختار این درخت مباحثی را مطرح کنیم اما پیش از آن، لازم است مباحث پیش نیاز لازم دوره شود.



شکل ۳: جامعه‌ای با  $2N = 10$  عضو و درخت Coalescence برای  $n = 4$  نمونه از این جامعه

### ۱.۳ توزیع هندسی

متغیر تصادفی  $T$  نشان دهنده تعداد دفعاتی است که لازم است یک آزمایش را تکرار کنیم تا به اولین پیروزی (رویدادی تعریف شده مثل ۱ (شیر) آمدن در پرتاب سکه) در آن آزمایش برسیم. به توزیع احتمال این متغیر تصادفی، توزیع هندسی گفته میشود که به صورت  $Geo(p)$  نمایش داده میشود که  $p$  پارامتر این توزیع و برابر با احتمال پیروزی است. این احتمال برابر است با:

$$\mathbb{P}[T = j] = (1 - p)^{j-1} p$$

همچنین داریم:

$$\mathbb{E}[T] = \frac{1}{p}$$

$$Var[T] = \frac{1-p}{p^2}$$

این توزیع، بی‌حافظه (*Memory Less*) است، به این معنی که:

$$\mathbb{P}[T > t_2 | T > t_1] = \mathbb{P}[T > t_2 - t_1]$$

یعنی به شرط اینکه داشته باشیم تا  $t_1$  پیروزی حاصل نشده است، احتمال اینکه تا  $t_2$  هم پیروزی حاصل نشود، مستقل است از این که قبلاً به مدت  $t_1$  به پیروزی نرسیده‌ایم و آن احتمال صرفاً برابر است با احتمال اینکه به اندازه تفاسیل  $t_1$  از  $t_2$  پیروزی حاصل نشود.

قضیه ۲. فرض کنیم  $S$  و  $T$  دو توزیع هندسی مستقل به صورت زیر باشند

$$T \sim Geo(p)$$

$$S \sim Geo(p')$$

آنگاه میتوان نوشت:

$$\min(T, S) = Geo(p + p' - pp')$$

اثبات. میدانیم که متغیر تصادفی  $T$  و  $S$  رویداد هایی را مدل میکنند که احتمال پیروزی در آنها به ترتیب برابر با  $p$  و  $p'$  است. میدانیم مقدار  $\min(T, S)$  به این معنی است که چندبار باید منتظر بمانیم تا در حداقل یکی از دو متغیر تصادفی  $S$  و  $T$  شاهد پیروزی باشیم. از طرفی احتمال اینکه در یک مرحله هر دوی متغیرها پیروزی نداشته باشند، برابر است با  $(1-p)(1-p')$ . بنابراین احتمال اینکه حداقل یکی از این دو متغیر پیروزی داشته باشند برابر است با  $1 - (1-p)(1-p')$  که مساوی است با  $p + p' - pp'$  که این همان صورت قضیه است.

### ۲.۳ توزیع نمایی

فرض کنیم متغیر تصادفی  $U$  دارای توزیع نمایی با پارامتر  $a$  باشد؛ آنگاه تابع توزیع احتمال این متغیر تصادفی برابر است با:

$$\mathbb{P}[U \leq t] = F_U(t) = 1 - e^{-at}$$

و تابع چگالی احتمال آن برابر است با:

$$f_U(t) = \begin{cases} ae^{-at} & t \geq 0 \\ 0 & o.w. \end{cases}$$

همچنین برای این متغیر تصادفی داریم:

$$\mathbb{P}[U > t_2 | U > t_1] = \mathbb{P}[U > t_2 - t_1]$$

بنابراین توزیع نمایی نیز بی‌حافظه است.

قضیه ۳. اگر  $U$  و  $V$  دو متغیر تصادفی با توزیع هندسی با پارامترهای  $a$  و  $b$  باشند، داریم:

$$\min(U, V) \sim \text{Exp}(a + b)$$

ارتباط دو توزیع. فرض کنیم  $T$  متغیر تصادفی زمان انتظار برای اولین آمدن سکه باشد که دارای توزیع هندسی است. میدانیم:

$$\begin{aligned} \mathbb{P}[T > j] &= \sum_{i=j+1}^{\infty} (1-p)^{i-1} p \\ &= p \sum_{i=j+1}^{\infty} (1-p)^{i-1} \quad (\text{دنباله هندسی}) \\ &= \frac{p(1-p)^j}{1-(1-p)} \\ &= (1-p)^j \end{aligned} \quad (2)$$

حال فرض کنیم سکه‌ای را با بازه‌های زمانی به اندازه  $M$  پشت سر هم پرتاب می‌کنیم و  $u$  را زمانی در نظر می‌گیریم که مقدار ۱ در پرتاب سکه حاصل شود. بنابراین با استفاده از رابطه (۲)، احتمال اینکه  $u$  بیشتر از مقدار زمان  $t$  باشد، برابر است با:

$$\mathbb{P}[u > t] = \mathbb{P}\left[\frac{u}{M} > \frac{t}{M}\right] = (1-p)^{\frac{t}{M}} \quad (3)$$

که در این رابطه  $u$  و  $t$  را هردو تقسیم بر  $M$  کردیم تا جنس حاصل تقسیم از نوع تعداد باشد تا بتوان از رابطه (۲) برای محاسبه احتمال استفاده کرد. حال اگر فرض کنیم  $M$  و  $P$  کوچک باشند اما  $\frac{P}{M} = a$ ، آنگاه با توجه به اینکه طبق فرض‌ها،  $1-p \approx e^{-p}$ ، بنابراین میتوان نوشت:

$$(1-p)^{\frac{t}{M}} \approx e^{-at} \quad (4)$$

بنابراین طبق (۳) و (۴) خواهیم داشت:

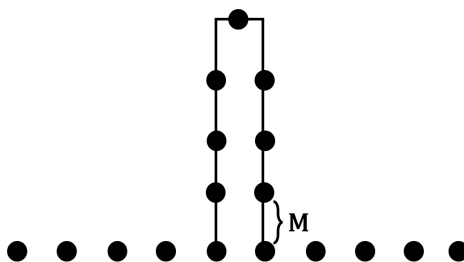
$$\mathbb{P}[u \leq t] = 1 - e^{-at}$$

که این همان توزیع نمایی است، بنابراین  $u$  دارای توزیع نمایی است. این بدین معنی است که اگر در توزیع هندسی اولیه  $T$ ، سکه را با سرعت خیلی زیاد و پشت سر هم پرتاب کنیم و احتمال ۱ آمدن هم مقدار خیلی کمی داشته باشد، آنگاه زمان انتظار برای ۱ آمدن تقریباً پیوسته شده و توزیع احتمال بدست آمده از نوع نمایی خواهد بود. در ادامه، از پیش‌نیازهای مطرح شده می‌خواهیم برای مدل کردن درخت Coalescent استفاده کنیم.

### ۳.۳ درخت Coalescent

فرض کنیم جامعه‌ای به اندازه  $2N$  داریم که در شرایط مدل Wright-Fisher صدق میکند. آخرین نسل این جمعیت را در نظر می‌گیریم و دو عضو از آن را انتخاب کنیم. مقدار زمانی که بین هر دو نسل وجود دارد را با  $M$  نشان می‌دهیم. شکل ۴ نمونه‌ای از این جامعه را نشان می‌دهد. تعداد نسل‌های مورد نیاز که باید به عقب برگردیم تا اینکه این دو عضو جد مشترک پیدا کنند را با متغیر تصادفی  $T$  نشان می‌دهیم. میدانیم احتمال جد مشترک پیدا کردن دو عضو در هر بار برگشت به نسل قبل برابر است با  $p = \frac{1}{2N}$ . حال برای اینکه جد مشترک این دو عضو را پیدا کنیم، باید تعدادی نسل انتظار یکشیم تا در نهایت این دو عضو جد مشترک پیدا بکنند. این انتظار از جنس انتظارهایی است که در توزیع هندسی و نمایی برای رسیدن به پیروزی داشتیم. بنابراین احتمال اینکه "باید تعداد  $j$  نسل انتظار بکشیم تا به جد مشترک این دو عضو برسیم" را به صورت زیر نشان می‌دهیم:

$$\mathbb{P}[T = j] = (1-p)^{j-1} p$$



شکل ۴: جامعه‌ای با فاصله  $M$  بین هر دو نسل آن

حال اگر بخواهیم مانند قبل، این توزیع احتمال را پیوسته کنیم، باید  $p$  و  $M$  را به سمت صفر ببریم (به این معنی که مقدار  $N$  بزرگ است) و فاصله بین دو نسل را نیز کوتاه در نظر بگیریم و فرض کنیم  $\frac{p}{M} = a$ . در این صورت زمان انتظار برای جد مشترک پیدا کردن دو عضو جامعه را میتوان به صورت توزیع نمایی با پارامتر  $a$  به صورت زیر نوشت:

$$a = \frac{1}{2NM} \quad (5)$$

همچنین تابع چگالی احتمال آن برابر خواهد بود با:

$$f_T(t) = \begin{cases} ae^{-at} & t \geq 0 \\ 0 & o.w. \end{cases}$$

به عنوان مثال میتوان فرض کرد که هر واحد  $t$  به اندازه ۱۰۰۰ سال است و فاصله بین هر دو نسل یعنی  $M$  برابر با ۲۰ باشد، بنابراین اگر  $a$  را ۱ در نظر بگیریم، داریم:

$$\frac{1}{2NM} = a \Rightarrow 2N = \frac{1}{M} = \frac{1000}{20} = 50$$

یعنی جمعیت جامعه برابر با ۵۰ خواهد بود. بنابراین اگر ۵۰ نفر داشته باشیم که هر ۲۰ سال نسل آنها تغییر کند و جمعیت همان ۵۰ بماند، در ۵۰ نسل (۱ واحد  $t = 1000 = 50 * 20$ ) یا بیشتر میتوان از رابطه تخمینی‌ای که برای تابع چگالی احتمال مطرح شد، استفاده کرد. بنابراین نمیتوان از روابط مطرح شده برای تعداد نسل کم به نسبت ۵۰ استفاده کرد و در این صورت باید از همان روابط گسسته اولیه استفاده کرد. طبق قسمت قبل، باید یک زمان نمایی با نرخ  $a$  منتظر بمانیم تا دو عضو انتخاب شده، والد مشترک پیدا کنند. حال اگر بتوان این زمان را برای سه عضو جامعه بدست آورد، میتوان برای کل جامعه این زمان را تعمیم داد. برای سه عضو از جامعه، مقدار  $1 - p$  (احتمال اینکه در نسل قبل هیچ کدام از این سه عضو والد مشترک پیدا نکرده باشند) برابر است با:

$$\begin{aligned} 1 - p &= \left( \frac{2N - 1}{2N} \right) \left( \frac{2N - 2}{2N} \right) \\ &= \left( 1 - \frac{1}{2N} \right) \left( 1 - \frac{2}{2N} \right) \\ &= 1 - \frac{1}{2N} - \frac{2}{2N} + \frac{2}{2N^2} \end{aligned}$$

با صرف نظر از عبارت آخر در مقابل با عبارت دوم و سوم (به این معنی که احتمال اینکه این سه عضو با هم در یک نسل والد مشترک پیدا کنند را قابل چشم پوشی در نظر گرفتیم)، میتوان نوشت:

$$p = \frac{3}{2N}$$

با داشتن  $p$ ، اگر فاصله میان دو نسل را برابر با  $M$  در نظر بگیریم، میتوان  $a$  (پارامتر توزیع نمایی) را به صورت زیر محاسبه کرد:

$$a = \frac{p}{M} = \frac{3}{2NM}$$

بنابراین برای اینکه دو عضو از سه عضو انتخابی والد مشترک پیدا کنند، باید مطابق با یک توزیع نمایی با پارامتر  $a = \frac{3}{2NM}$  منتظر بمانیم. پس از والد مشترک پیدا کردن دو عضو از سه تا، از آنجا که سیستم بدون حافظه است، دوباره زمان انتظار والد مشترک پیدا کردن دو عضو باقیمانده از نوع نمایی اما این بار با نرخ  $a = \frac{1}{2NM}$  (رابطه (۵)) خواهد بود.

بنابراین برای اینکه از سه عضو به عضو والد مشترک همه برسیم، باید در ابتدا به صورت نمایی با نرخ  $\frac{3}{2NM}$  و بعد از اولین والد مشترک پیدا کردن، به صورت نمایی با نرخ  $\frac{1}{2NM}$  منتظر بمانیم. حال میتوان در حالت کلی، درخت Kingsman را تعریف کرد.

### ۴.۳ درخت Kingsman

فرض کنیم  $n$  عضو در جامعه داریم. برای اینکه به  $MRCA$  (Most Recent Common Ancestor) که همان جد مشترک تمام  $n$  عضو است برسیم، باید زمان‌هایی از جنس نمایی برحسب تعداد نسل منتظر بمانیم. توجه کنیم که با والد مشترک پیدا کردن هر دو عضو، پارامتر توزیع نمایی تغییر پیدا میکند. این قدر منتظر میمانیم تا تمام اعضا به صورت دو به دو با یکدیگر والد مشترک پیدا کنند و در نهایت به  $MRCA$  برسیم. اگر بخواهیم نرخ توزیع نمایی را برای  $k$  عضو بدست بیاوریم، ابتدا باید مقدار احتمال والد مشترک پیدا کردن در نسل قبلی را برای  $k$  عضو بدست بیاوریم:

$$\begin{aligned} 1 - p &= \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \dots \left(1 - \frac{k-1}{2N}\right) \\ &\approx 1 - \frac{1 + 2 + 3 + \dots + k-1}{2N} \\ &= 1 - \frac{k(k-1)}{2 \times 2N} \\ &= 1 - \frac{\binom{k}{2}}{2N} \end{aligned}$$

(در تساوی دوم از توان‌های بالای  $N$  در مخرج صرف نظر شده است)

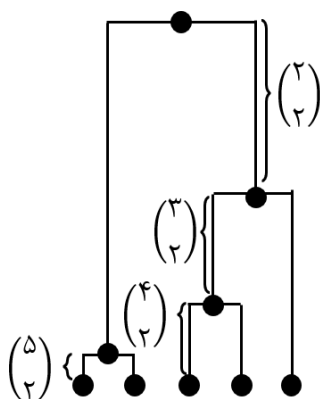
بنابراین:

$$p = \frac{\binom{k}{2}}{2N}$$

بنابراین پارامتر توزیع نمایی برای  $k$  عضو با فرض اینکه  $M = \frac{1}{2N}$  برابر است با:

$$a = \frac{p}{M} = \frac{\binom{k}{2}}{2NM} = \binom{k}{2}$$

میدانیم امیدریاضی توزیع نمایی برابر است با  $\frac{1}{a}$ .  $E[u] = \frac{1}{a}$ . بنابراین هرچه مقدار نرخ  $a$  افزایش پیدا کند، متوسط زمان انتظار برای پیروزی نیز کمتر خواهد بود. به عنوان مثال برای ۵ عضو میتوان پارامترهای توزیع نمایی در یک درخت *Kingsman Coalescent* نمونه را به صورت شکل ۵ در نظر گرفت.



شکل ۵: نمایش نرخ توزیع نمایی برای هر والد مشترک پیدا کردن در یک درخت Kingsman Coalescent نمونه

حال که توانستیم درباره طول یال‌های درخت Coalescent کسب کنیم، می‌خواهیم ببینیم چطور میتوان ساختار این درخت را بدست آورد. برای اینکار نیاز به اطلاعات جهش‌های اتفاق افتاده در ژنوم این اعضا داریم. نحوه بدست آوردن درخت Coalescent با استفاده از اطلاعات جهش در جلسات آینده مطرح خواهد شد.