



# ژنومیک محاسباتی

مطهری و فروغمند  
پاییز ۱۴۰۰

## بررسی مدل‌های تولید مثل و درخت ادغام به همراه جهش

جلسه سیزدهم

نگارنده: علی قاری‌زاده

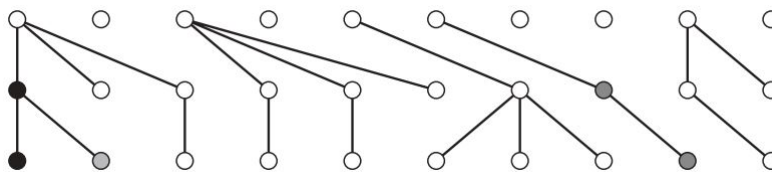
### ۱ مروری بر مباحث گذشته

جلسه قبل در ابتدا در مورد مدل جمعیت رایت-فیشر صحبت شد. این مدل، یک مدل پایه برای تولید مثل است و توصیفی پویا از تکامل یک جمعیت ایده‌آل و انتقال ژن‌ها از نسلی به نسل دیگر را ارائه می‌کند [HSW05]. سپس با توضیح توزیع‌های هندسی و نمایی و بررسی برخی از ویژگی‌های آن‌ها، مثل ویژگی بی‌حافظه بودن، فرآیند ادغام کینگمن مورد بررسی قرار گرفت. نهایتاً به این نتیجه رسیدیم که در زمان پیوسته، برای  $k$  فرد می‌توانیم زمان ادغام را با توزیع نمایی با پارامتر  $(\frac{k}{2})$  تخمین بزنیم.

### ۲ مدل رایت-فیشر با جهش

در مباحثی که در جلسات قبل مطرح می‌شد، هنگام تولید مثل فرض می‌کردیم هیچ جهشی رخ نمی‌دهد یعنی ژنوتایپ همه‌ی فرزندان مشابه والدشان بود. در این بخش می‌خواهیم وجود جهش را هم به مدل قبلی اضافه کنیم. در این مدل فرض می‌کنیم هر ژن به احتمال  $u$  هنگام انتقال به نسل دیگر دچار جهش می‌شود. همچنین فرض می‌کنیم که جهش‌ها خنثی هستند یعنی احتمال اتفاق افتادن یک جهش در همه‌ی انواع ژن‌ها برابر است. شکل ۱ یک مدل رایت-فیشر با جهش را نشان می‌دهد. که از نسل اول به دوم ۲ جهش و از نسل دوم به سوم ۱ جهش در جمعیت رخ داده است. در جلسات قبل گفته شد که اگر در یک جمعیت با اندازه  $2N$  که همه‌ی آن‌ها ژن  $A$  را دارند یکی از آن‌ها جهش پیدا کند و ژن  $a$  را به دست آورد با احتمال  $\frac{1}{2N}$  در آینده ژن  $a$  در بین تمام جمعیت تثبیت می‌شود و با احتمال  $\frac{2N-1}{2N}$  از بین جمعیت رانده می‌شود و از بین می‌رود. این به این معناست که یک جهش خنثی به احتمال کمی در جامعه باقی می‌ماند. اما اگر فرض کنیم که موجودات جامعه تعداد زیادی ژن در کنار هم دارند و در یک ژنوم  $2N$  جهش رخ داده است امید ریاضی ما برای تعداد جهش‌های خنثی که در جامعه تثبیت می‌شوند برابر ۱ خواهد بود. پس جهش‌های خنثی هم

شانس باقی ماندن در جامعه را دارند. و یکی از مباحث جدی در ژنتیک جمعیت همین مسئله است که وقتی می‌بینیم یک مکان از ژنوم در یک جامعه تثبیت شده است بتوانیم بفهمیم که به خاطر انتخاب طبیعی این اتفاق افتاده است یا اینکه مشابه مثال بالا یک جهش خنثی بوده که تثبیت شده است.

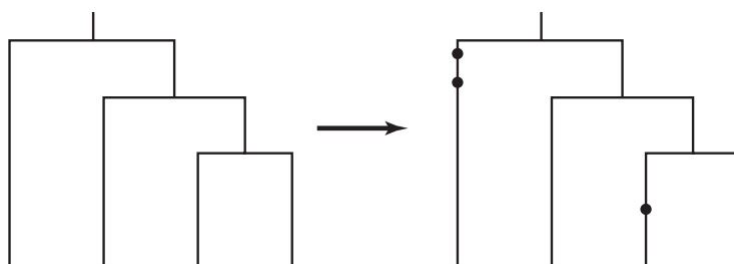


شکل ۱: مدل رایت-فیشر با جهش [HSW05]

### ۳ ساخت درخت ادغام با در نظر گرفتن جهش

ما قبلاً با استفاده از مدل کینگمن روشی برای ساخت درخت ادغام ارائه کردیم. اکنون می‌خواهیم مدلی بسازیم که جهش‌ها را هم در درخت ادغام وارد کنیم. در این مدل فرض می‌کنیم که جهش‌ها مستقل از هم رخ می‌دهند. از دو منظر میتوان این مسئله را بررسی کرد:

۱: منظر اول به این شکل است که درخت ادغام را به صورت مستقل می‌سازد و سپس تعداد جهش‌های موجود در هر شاخه را تخمین می‌زند. احتمال اتفاق افتادن  $m$  جهش در یک شاخه که در آن  $k$  نسل عوض شده است برابر  $(1-p)^{k-m} p^m \binom{m}{k}$  است. میدانیم اگر  $p$  به صفر میل کند و  $k$  به بی‌نهایت میل کند و  $pk$  به  $\theta$  میل کند. توزیع پواسون با پارامتر  $\theta$  تقریب خوبی از توزیع دوجمله‌ای بالا خواهد بود. پس تعداد جهش‌های هر شاخه را میتوانیم با توزیع پواسون به دست آوریم. شکل ۲ ساخت درخت با این روش را نشان می‌دهد که ابتدا درخت ادغام ساخته شده و سپس تعداد جهش‌های روی هر شاخه محاسبه می‌شود.



شکل ۲: ساخت درخت ادغام با جهش از منظر یک [HSW05]

۲: منظر دوم به این شکل است که وقتی در زمان به عقب برمی‌گردیم یا جهش می‌بینیم یا ادغام دو موجود. زمانی که باید صبر کنیم در بین  $k$  موجود اولین جهش را ببینیم از توزیع نمایی با پارامتر  $\frac{k\theta}{4}$  است (این با استفاده از خاصیت مینوم توزیع‌های نمایی که در جلسه قبل بررسی شد به دست می‌آید). همچنین زمانی که باید صبر کنیم تا یک ادغام هم ببینیم از توزیع نمایی با پارامتر  $\binom{k}{2}$  به دست می‌آید. پس زمان انتظار ما برای دیدن یک جهش یا ادغام از توزیع نمایی با پارامتر  $\frac{k\theta}{4} + \frac{k(k-1)}{4}$  است. پس حالا میدانیم در یک زمانی جهش یا ادغام رخ داده است. سوال این است که کدامیک رخ داده است؟ حدث خوب این است که بگوییم با احتمال  $\frac{\frac{k\theta}{4}}{\binom{k}{2} + \frac{k\theta}{4}}$  جهش رخ داده و با احتمال  $\frac{\binom{k}{2}}{\binom{k}{2} + \frac{k\theta}{4}}$  ادغام رخ داده است. شکل ۳ ساخت درخت با این روش را نشان می‌دهد که در آن ابتدا زمان‌های رخ دادن یک اتفاق مشخص شده است (خطچین‌ها). سپس خود اتفاق با احتمالات گفته شده انتخاب شده است.

برای اثبات اینکه ۲ احتمال بالا درست هستند باید ثابت کنیم که اگر  $X$  متغیر تصادفی با توزیع نمایی با پارامتر  $\lambda_1$  باشد و  $Y$  توزیع نمایی با پارامتر  $\lambda_2$  باشد. آنگاه  $P[X \leq Y] = \frac{\lambda_1}{\lambda_1 + \lambda_2}$  خواهد بود.

اثبات.

$$\begin{aligned} P[X \leq Y] &= \int_{y=0}^{\infty} \int_{x=0}^y f_{xy}(x, y) dx dy \\ &= \int_{y=0}^{\infty} \int_{x=0}^y f_x(x) f_y(y) dx dy \\ &= \int_{y=0}^{\infty} \int_{x=0}^y \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 y} dx dy \end{aligned}$$

$$\begin{aligned}
 &= \int_{y=0}^{\infty} \lambda_2 e^{-\lambda_2 y} \left[ \int_{x=0}^y \lambda_1 e^{-\lambda_1 x} dx \right] dy \\
 &= \int_{y=0}^{\infty} \lambda_2 e^{-\lambda_2 y} [1 - e^{-\lambda_1 y}] dy
 \end{aligned}$$

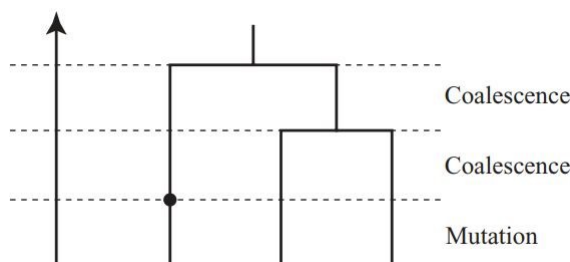
اگر قسمت بیرون کروش را در داخل کروش ضرب کنیم دو انتگرال به دست می‌آید که انتگرال سمت چپ برابر با انتگرال از یک توزیع نمایی است و مقدار آن برابر یک خواهد بود.

$$\begin{aligned}
 &= 1 - \int_0^{\infty} \lambda_2 e^{-(\lambda_1 + \lambda_2)y} dy \\
 &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \int_0^{\infty} (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)y} dy
 \end{aligned}$$

انتگرال بالا انتگرال توزیع نمایی است که مقدار آن برابر یک خواهد بود.

$$\begin{aligned}
 &= 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \\
 &= \frac{\lambda_1}{\lambda_1 + \lambda_2}
 \end{aligned}$$

□



شکل ۳: ساخت درخت ادغام با جهش از منظر دو [HSW05]

## ۴ تخمین گر واترسون

قضیه واترسون توزیع احتمال تعداد جهش‌های موجود برای جمعیتی با اندازه  $n$  را به شکل زیر به دست می‌آورد:

$$P[S_n = s] = \frac{n-1}{\theta} \sum_{i=1}^{n-1} (-1)^{i-1} \binom{n-1}{i-1} \left( \frac{\theta}{i+\theta} \right)^{s+1}$$

که در آن  $S_n$  تعداد کل جهش‌های موجود و  $\theta$  نرخ جهش است. حال می‌خواهیم امید ریاضی  $S_n$  را محاسبه کنیم. ابتدا از منظر اول به مسئله نگاه میکنیم. تعداد کل جهش‌های رخ داده در تمام جمعیت از زمان حال تا زمان رخ دادن اولین ادغام را  $X_n$  می‌نامیم. به طور مشابه تعداد جهش‌های رخ داده از زمان اولین ادغام تا دومین ادغام را  $X_{n-1}$  می‌نامیم. به همین شکل می‌توانیم تا  $X_2$  تعریف کنیم. در این حالت داریم:

$$S_n = X_n + X_{n-1} + \dots + X_2$$

$$\Rightarrow E[S_n] = E[X_n] + E[X_{n-1}] + \dots + E[X_2]$$

اگر زمان اتفاق افتادن اولین ادغام برای  $k$  نفر را  $t$  بنامیم داریم:

$$E[X_k] = E_t E[X_k | t]$$

و میدانیم که تعداد جهش‌های یک فرد از توزیع پواسون تبعیت می‌کند و امید ریاضی توزیع پواسون با پارامتر  $\lambda$  خود  $\lambda$  است. بنابراین:

$$E[X_k] = \lambda k E[t]$$

توزیع  $t$  هم‌نمایی است و امید ریاضی توزیع نمایی با پارامتر  $\lambda$  برابر  $\frac{1}{\lambda}$  است. و پارامتر توزیع نمایی برای زمان انتظار اولین ادغام برابر  $\binom{k}{2}$  است. بنابراین:

$$E[X_k] = \frac{2Nk}{\binom{k}{2}} \\ = \frac{4N}{k-1}$$

همچنین گفته بودیم که  $\frac{\theta}{4} = 2Nx$  پس  $\theta = 4Nx$  و اگر واحد زمان  $x$  را برابر ۱ بگیریم داریم:

$$\theta = 4N$$

بنابراین:

$$E[X_k] = \frac{\theta}{k-1}$$

همچنین در قسمت قبل گفتیم:

$$E[S_n] = E[X_n] + E[X_{n-1}] + \dots + E[X_2] \\ \Rightarrow E[S_n] = \frac{\theta}{n-1} + \frac{\theta}{n-2} + \dots + \frac{\theta}{1} \\ \Rightarrow E[S_n] = \theta \sum_{i=1}^{n-1} \frac{1}{i}$$

اگر داشته باشیم  $h_n = \sum_{i=1}^{n-1} \frac{1}{i}$  آنگاه:

$$E[S_n] = \theta h_n$$

بنابراین اگر ما تخمینی از تعداد کل جهش‌ها داشته باشیم می‌توانیم تخمینی از نرخ جهش را به دست آوریم. نرخ جهش واترسون:

$$\theta_W = \frac{S_n}{h_n}$$

می‌توان واریانس  $S_n$  را نیز به مشابه با روش بالا به دست آورد که در این حالت داریم:

$$Var[S_n] = \theta h_n + \theta^2 \sum_{i=1}^{n-1} \frac{1}{i^2}$$

که اگر  $\sum_{i=1}^{n-1} \frac{1}{i^2}$  را  $g_n$  بنامیم داریم:

$$Var[S_n] = \theta h_n + \theta^2 g_n$$

## ۵ به دست آوردن توزیع احتمال تعداد جهش‌ها از منظر دوم

اگر بخواهیم  $P[S_n = s]$  را از منظر دوم به دست بیاوریم داریم:

$$P[S_n = s] = P[S_n = s | \text{اول جهش ببینیم}] P[\text{اول جهش ببینیم}] + P[S_n = s | \text{اول ادغام ببینیم}] P[\text{اول ادغام ببینیم}] \\ = P[S_n = s | \text{اول جهش ببینیم}] \frac{\frac{n\theta}{2} + \frac{n\theta}{2}}{\binom{n}{2} + \frac{n\theta}{2}} + P[S_n = s | \text{اول ادغام ببینیم}] \frac{\binom{n}{2}}{\binom{n}{2} + \frac{n\theta}{2}} \\ = P[S_n = s-1] \frac{\frac{n\theta}{2} + \frac{n\theta}{2}}{\binom{n}{2} + \frac{n\theta}{2}} + P[S_{n-1} = s] \frac{\binom{n}{2}}{\binom{n}{2} + \frac{n\theta}{2}}$$

که این یک رابطه بازگشتی به ما می‌دهد.

## ۶ به دست آوردن توزیع احتمال تعداد جهش‌ها از منظر اول

اگر جمع طول همه‌ی شاخه‌های درخت ادغام را برابر  $t$  قرار دهیم یعنی  $t = nt_1 + (n-1)t_2 + \dots$  که در آن  $t_1$  و  $t_2$  و ... از توزیع‌های نمایی با پارامترهای  $(\frac{n}{2})$  و  $(\frac{n-1}{2})$  و ... هستند، داریم:

$$P[S_n = s] = \int_0^\infty P[S_n = s|t]f(t) dt$$

که در آن  $P[S_n = s|t] = \text{Poisson}(\frac{\theta}{2}) dt$  و  $f(t)$  توزیع احتمال روی  $t$  است. انتگرال بالا همان عبارت واترسون را که در ابتدای بخش ۴ گفته شد، نتیجه می‌دهد.

## ۷ مقدمه جلسه بعد

تا این قسمت درس فرض می‌کردیم که فقط یک مکان در ژنوم هر فرد وجود دارد. جلسه بعد می‌خواهیم دیدگاهمان را از تک مکان به چند مکان عوض کنیم البته بدون در نظر گرفتن نوترکیبی. در این حالت به دلیل در نظر نگرفتن نوترکیبی ساختار درخت ادغام با درخت فعلی تفاوتی نخواهد کرد. چون در حالتی که نوترکیبی وجود دارد یک فرزند در درخت از دو والد نشأت می‌گیرد اما در حالتی که ما بررسی می‌کنیم هر فرزند از یک والد می‌آید. رخ دادن جهش‌ها هم باز روی درخت نشان داده می‌شود. اما اثر جهش‌ها روی توالی‌ها متفاوت خواهد بود چون هر جهش در هر قسمت از درخت می‌تواند مکان خاصی از ژنوم را عوض کند. برای بررسی این حالت نیاز داریم تا مدل‌های جهش را تعریف کنیم.

### ۱.۷ مدل‌های جهش

- ۱: مدل بی‌نهایت مکان: در این مدل فرض می‌شود که طول ژنوم بی‌نهایت است و هر جهش در یک جای جدید از ژنوم اتفاق می‌افتد. فقط در این مدل است که ما تمام جهش‌های اتفاق افتاده در طول زمان را می‌توانیم ببینیم.
- ۲: مدل محدود مکان: در این مدل امکان اتفاق افتادن دو جهش در یک مکان وجود دارد.
- ۳: مدل بی‌نهایت آلل: در این مدل به مکان جهش کاری نداریم و فرض می‌کنیم که هر جهشی که اتفاق می‌افتد حتماً یک آلل جدید ایجاد می‌شود.

## مراجع

[HSW05] Jotun Hein, Mikkel H. Schierup, and Carsten Wiuf. *Gene genealogies, variation and evolution*. Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2005.