



ژنومیک محاسباتی

مطهری و فروغمند
پاییز ۱۴۰۰

ژنتیک جمعیت

جلسه یازدهم

نگارنده: مرجان باغ گلشنی

۱. مروری بر مباحث گذشته

در ده جلسه گذشته در مورد چستی درخت تبارزایی، چگونگی بازسازی آن و الگوریتم های موجود به تفصیل صحبت کردیم. از این جلسه وارد بخش دوم درس که ژنتیک جمعیت هست می شویم.

۲. ژنتیک جمعیت

در این درس هدف این است که یک جمعیت را از لحاظ ژنتیکی مطالعه کنیم. می دانیم هر فرد یک ژنوم دارد و ژنوم های افراد با یکدیگر متفاوت است، مثلاً قومیت ایرانی و اروپایی دارای ژنوم های متفاوت هستند، اما دلیل این تفاوت در ژنوم چیست؟ (برای سادگی در این درس ما از کل ژنوم فقط یک location را در نظر میگیریم).

اگر یک نقطه مشترک را بر روی ژنوم همه انسان ها در نظر بگیریم (مثلاً نوکلئوتاید شماره ۱۰) ممکن است حرف A یا C یا G یا T ببینیم. برای که به آنها الیل (allele) گفته می شود.

فرض کنید فقط دو الیل A و C را داشته باشیم، به یکی از آنها Major الیل و به دیگری Minor الیل می گویند. حال بررسی می کنیم و می بینیم که جمعیت اروپایی اکثراً دارای الیل A هستند ولی درصد بسیار کمی الیل C نیز در آنها مشاهده می شود، از طرفی در جمعیت ایرانی ها اکثراً دارای الیل C هستند اما درصد بسیار کمی با الیل A نیز در آنها وجود دارد، بررسی دلیل این تفاوت همان مبحث مورد مطالعه ژنتیک جمعیت است.

به طور کلی ما در ژنتیک جمعیت ۴ مورد عمده را بررسی می کنیم:

- جهش: فرض کنیم همه یک جد مشترک داشتند که در آنجا الیل A وجود داشته و دلیل اینکه الان ما الیل C را مشاهده می کنیم، جهش است. پس جهش یک عاملی است که ما فرکانس الیل ها را متفاوت می بینیم.

- انتخاب طبیعی: در یک الیل جهش به وجود آمده است و موجود در پی آن جهش یک قوتی را به دست آورده است و به دلیل آن قوت در جمعیت زیاد می‌شود.
 - رانش ژنتیکی: یعنی به صورت تصادفی یک جهش که نه قوتی به موجود اضافه میکند و نه ضعیفی ایجاد میکند به صورت تصادفی در جمعیت زیاد می‌شود. قابل ذکر است که رانش ژنتیکی مستقل از انتخاب طبیعی است؛ زیرا در انتخاب طبیعی قوتی به موجود افزود می‌شد اما در رانش ژنتیکی صرفاً به خاطر تصادفی بودن یک الیل تمایز به وجود می‌آید.
 - جریان ژن: هنگامی که یک موجود به یک مکان دیگر می‌رود و در آنجا زاد و ولد میکند و جریان آنجا را به هم میریزد. در واقع یک مترال ژنتیکی از یک قوم به یک قوم دیگر انتقال می‌یابد.
- پس به طور کلی مطالعه فرکانس الیلی در جمعیت و قومیت‌ها، پیدا کردن دلیل برای آنها و توجیه شان مبحث مورد بررسی در ژنتیک جمعیت است.

۳ قانون هاردی-وینبرگ

می‌دانیم موجودات از سه دسته هستند:

- Haploid : از هر کروموزوم یکی داریم.
 - Diploid : از هر کروموزوم دو تا داریم.
 - Moltyploid : از هر کروموزوم چندین تا داریم.
- قابل ذکر است، قانون هاردی وینبرگ را برای diploid مورد بررسی قرار می‌دهیم.
- فرض‌های زیر را برای ساده سازی انجام می‌دهیم.
- جمعیت diploid است.
 - ازدواج‌ها کاملاً تصادفی است.
 - نیروهای انتخاب طبیعی وجود ندارد.
 - جمعیت بی‌نهایت است.
 - نسل‌ها در هم نیستند (یک نسل فقط با نسل خود ازدواج می‌کند).

نسل i را در نظر بگیرید که شامل M مرد و W زن هستند. بحث ما در مورد یک نقطه از ژنوم است که آن نقطه از ژنوم شامل ۲ الیل است. اگر ۲ الیل یکسان باشند موجود را Homozygous می‌نامیم و اگر ۲ الیل متفاوت با یکدیگر باشند، موجود را Heterozygous می‌نامیم. حال فرض کنید:

- کسر X از نسل i Homozygous $[(0,0) \text{ or } (A,A)]$ هستند.
 - کسر Y از جمعیت نسل i Hetrozygous $[(0,1) \text{ or } (C,A)]$ هستند.
 - کسر Z از جمعیت نسل i Homozygous $[(1,1) \text{ or } (C,C)]$ هستند.
- سوالی که به وجود می‌آید این است که در نسل $i+1$ با چه احتمالی یک فرد از نوع $(0,0)$ است؟
- جواب:

$$P(i - th = (0,0)) =$$
$$P[(0,0)|m = (0,0) \text{ and } w = (0,0)] * P(m = (0,0) \text{ and } w = (0,0))$$

$$\begin{aligned}
& +P[(0,0)|m=(0,1)andw=(0,0)] * P(m=(0,1)andw=(0,0)) \\
& +P[(0,0)|m=(0,0)andw=(0,1)] * P(m=(0,0)andw=(0,1)) \\
& +P[(0,0)|m=(0,1)andw=(0,1)] * P(m=(0,1)andw=(0,1)) \\
& = 1 * X^2 + \frac{1}{2} * 2XY + \frac{1}{2} * 2XY + \frac{1}{4} * 4Y^2
\end{aligned}$$

می دانیم که برای اینکه در نسل ۱+۱ موجودی به صورت (۰،۰) باشد نمی تواند مرد یا زنی به صورت (۱،۱) والد باشد در واقع احتمال وجود والد به صورت (۱،۱) صفر است. به همین شکل احتمال را برای اینکه فرد از نوع (۱،۱) باشد نیز حساب می کنیم و داریم:

$$\begin{aligned}
& P[(1,1)|m=(1,1)andw=(1,1)] * P(m=(1,1)andw=(1,1)) \\
& +P[(1,1)|m=(0,1)andw=(1,1)] * P(m=(0,1)andw=(1,1)) \\
& +P[(1,1)|m=(1,1)andw=(0,1)] * P(m=(1,1)andw=(0,1)) \\
& +P[(1,1)|m=(0,1)andw=(0,1)] * P(m=(0,1)andw=(0,1)) \\
& = 1 * Z^2 + \frac{1}{2} * 2ZY + \frac{1}{2} * 2ZY + \frac{1}{4} * 4Y^2
\end{aligned}$$

مقدار $X+Y$ را q می نامیم و همچنین مقدار $Y+Z$ را P می نامیم؛ از طرفی می دانیم که $X + 2Y + Z = 1$ و از این نتیجه میگیریم $p+q$ نیز برابر یک است.

برای محاسبه مقدار احتمال (۰،۱)، مقدار ۲ احتمال محاسبه شده را از یک کم می کنیم و داریم:

$$\begin{aligned}
P[0,0] &= X^2 + 2XY + Y^2 = (X + Y)^2 = p^2 \\
P[1,1] &= Z^2 + 2ZY + Y^2 = (Z + Y)^2 = q^2 \\
P[0,1] &= 1p^2q^2 = (p + q)^2 - p^2q^2 = 2pq
\end{aligned}$$

به طور کلی اگر یک نقطه ای باشد که در شرایط ایده آل صدق کند، آنگاه برای فهمیدن اینکه هر کسر از جامعه به چه گونه است حتما داریم:

	(0,0)	(0,1)	(1,1)
—	X	$2Y$	Z
$i - th$	$(X + Y)^2 = p^2$	$2(X + Y)(Z + Y) = 2pq$	$(Y + Z)^2$
$i + 1 - th$	$(p^2 + pq)^2$	$2(p^2 + pq)(q^2 + pq)$	$(q^2 + pq)^2$
$i + 2 - th$			

از آنجایی که میدانیم $p+q$ برابر مقدار یک است، میتوانیم سطر آخر را به صورت زیر بازنویسی کنیم.

	(0,0)	(0,1)	(1,1)
—	p^2	$2pq$	q^2
$i + 2 - th$			

به طور خلاصه p را احتمال تک صفر بودن جمعیت و q را احتمال تک یک بودن جمعیت در نظر میگیریم، حال برای اینکه بفهمیم یک جمعیت ۱۰۰۰ نفره در تعادل هاردی-وینبرگ هست یا نه، در قدم اول تمام جمعیت را ژنوتایپینگ میکنیم، سپس تعداد ۰ ها را شمرده و بر ۲۰۰۰ تقسیم میکنیم (چون هر فرد ۲ الیل دارد)؛ برای یک ها نیز به همین صورت، باتوجه به تعادل هاردی وینبرگ همو (۰،۰) برابر p^2 و همو (۱،۱) برابر q^2 و هترو برابر $2pq$ است از طرفی دیگر ما تعداد همو و هترو را از ژنوتایپینگ به دست آورده ایم، حال اگر ۳ عدد به دست آمده با سه عدد محاسبه شده اختلاف کمی داشته باشند می توانیم بگوییم تعادل هاردی-وینبرگ برقرار است، نتیجه میگیریم در این نقطه شروط ذکر شده برقرار است.

۴ Genetic Drift

تا به اینجا در مورد ازدواج های تصادفی در مقیاس زیاد بدون سلکشن صحبت کردیم حال در نظر بگیرید جمعیت کم باشد، ازدواج ها تعداد کمی هستند و ممکن است پس از چندین نسل حتی بعضی از الیل ها از بین بروند، به طور کلی به بررسی فرکانس الیل ها در طول زمان برای یک جمعیت نسبتاً کوچک Genetic Drift گفته می شود.

چون جمعیت را کوچک در نظر گرفتیم دیگر مانند قبل کسر های جمعیت مربوط به هم و هترو بودن ثابت نمی ماند و این نسبت در طول زمان در حال تغییر است و ثابت میشود که پس از گذشت زمان زیادی، همه جامعه به سمت تک الیلی شدن پیش می روند و الیل دیگر از جمعیت رانش (Drift) میکند و تمام میشود.

نکته: در صورتی که جمعیت ما به یک تعادل برسد، دیگر drift برای یک الیل رخ نمیدهد، اما لازم به ذکر است در انجا نمیتوانیم از تعادلی هاردی-وینبرگ استفاده کنیم، زیرا فرض کردیم جمعیت ما کوچک و ثابت است در حالی که تعادلی هاردی-وینبرگ برای جمعیت های بزرگ برقرار است.

۵ مدل Wright-Fisher

فرض هایی که در این مدل داریم:

- مدل Wright-Fisher بر روی haploid مطالعه میشود.
- نسل ها مجزا هستند.
- جمعیت ثابت است.
- همه به یک اندازه fit هستند. (selection وجود ندارد)
- جمعیت ساختار جغرافیایی و اجتماعی ندارد.

رابطه نسلی یا genealogy: همه افراد در نسل $i+1$ دقیقاً یک والد از نسل i ام دارند اما در نسل i ام افرادی هستند که هیچ فرزندی نداشته باشند. سوال: یک فرزند چگونه والد خود را انتخاب میکند؟

جواب: از دیدگاه Wright-Fisher این اتفاق کاملاً تصادفی و مستقل است. یعنی اگر در نسل i ام جمعیت ما $2N$ باشد، احتمال اینکه هر فردی از نسل i ام یک والد باشد برای فرزند نسل $i+1$ ام برابر $1/2N$ است.

درخت coalescent: این درخت مدلی از توزیع واگرایی الیل ها در یک genealogy است که به طور گسترده ای برای تخمین ژنتیکی جمعیت استفاده می شود، این درخت مستقل از جهش و فرکانس الیل ها است و روابط فامیلی را نشان می دهد؛ همچنین به ریشه آن MRCA میگویند. به زبان ساده تر یعنی اگر تمام الیل ها را به صورت انفرادی در نظر بگیریم و در زمان به عقب برگردیم برای آن ها یک جد مشترک پیدا خواهیم کرد، چون به یک جد مشترک میرسیم پس متوجه میشویم که drift رخ داده است و هنگامی میتوانیم بگوییم drift رخ نداده که به MRCA برسیم.

۶ مباحث آینده

در این جلسه سعی شد بیشتر به تعاریف اولیه و مبحث کلی که قرار است در ده جلسه آینده تدریس شود، بپردازیم. در جلسات آینده به تفصیل به بررسی دو مدلی که گفته شد میپردازیم.



۷ ارجاع و منابع

مراجع

برای این مبحث مرجعی معرفی نشده است.