An Overview of the Luria-Delbrück Distribution

Qi Zheng

School of Rural Public Health
Texas A&M Health Science Center
College Station, Texas

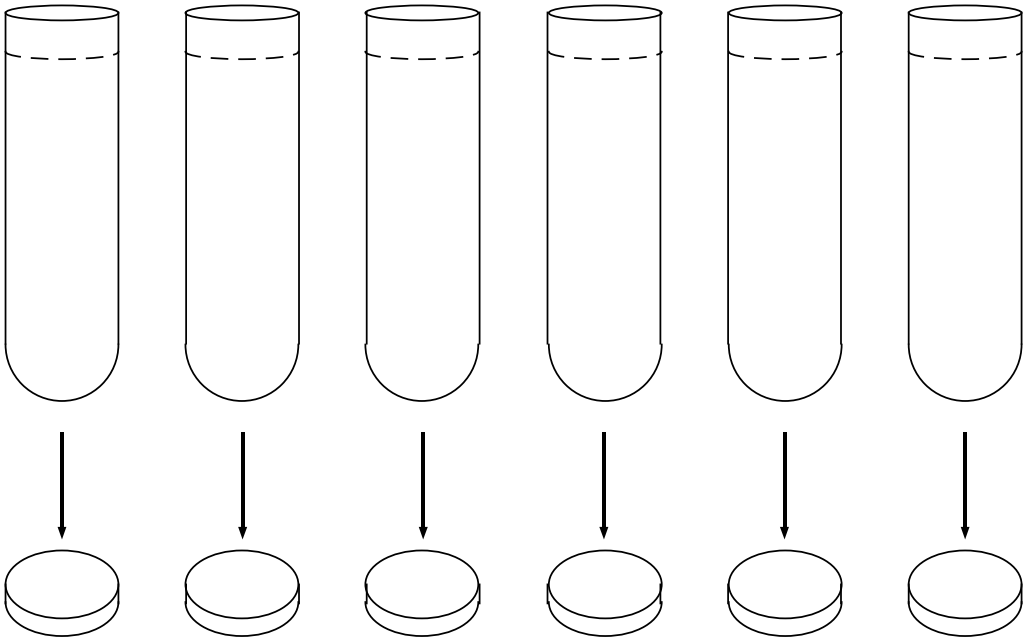
¶Where did the Luria-Delbrück distribution come from?

- in the 1940s, bacteria were believed to be different
- crucial issue: is bacterial mutation pre-adaptive or post-adaptive

- Luria (Watson's advisor) was preoccupied with this issue
- A solution was conceived at a faculty dance at Indiana Univ, while Luria was watching a slot machine
- see page 75 of *A Slot Machine, A Broken Test Tube*

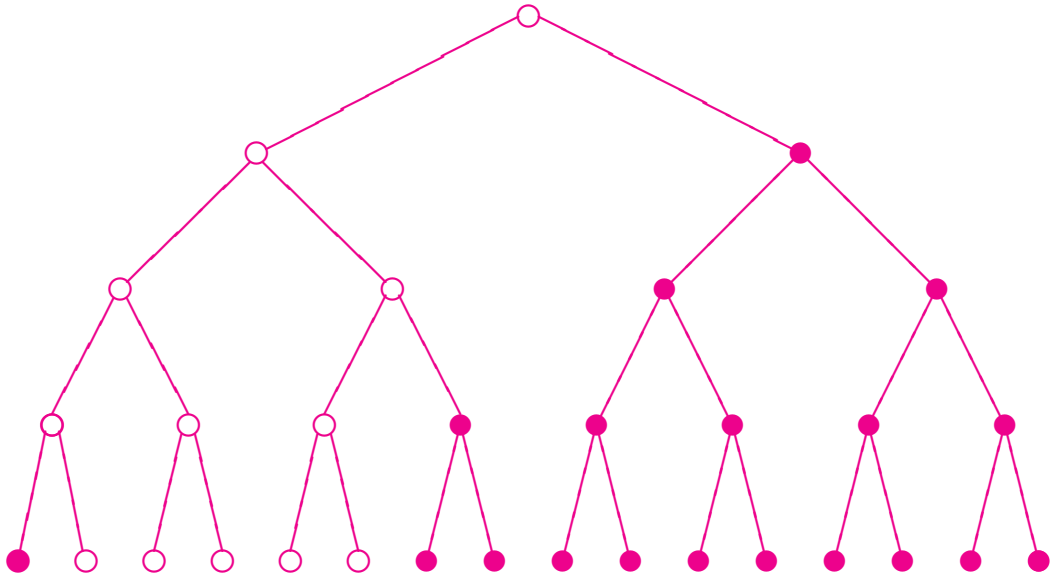
¶What is a fluctuation experiment?

- let bacteria grow in a liquid culture (incubation)
- transfer the contents of a tube to a solid culture (plating)



¶no models, no estimation of mutation rates

- first model proposed by Luria and Delbrück (1943), and modified by Lea and Coulson (1949)
- But the following illustrates the salient features



¶What's the major obstacle?

- p.g.f for the L-C model (1949)

$$G(z; m, \phi) = \exp\left\{\frac{m}{\phi}\left(\frac{1}{z} - 1\right)\log(1 - \phi z)\right\}$$

where $\phi = 1 - e^{-\beta T} < 1$ with β denoting cellular birth rate.

- Ma. et al. (1993) improved the L-C method, proposing a recursive algorithm

$$p(0; m, \phi) = e^{-m}$$

$$p(k; m, \phi) = \frac{m}{k}\sum_{j=1}^{k}\phi^{j-1}\left(1 - \frac{j\phi}{j+1}\right)p(k-j; m, \phi) \qquad (k \geq 1)$$

- How to make point and interval estimation of m?

¶A solution came rather unexpectedly (2005)

- Just differentiate the p.g.f.!!

$$\frac{\partial^i G}{\partial m^i} = \left[\frac{1}{\phi}\left(\frac{1}{z}-1\right)\log(1-\phi z)\right]^i \exp\left[\frac{m}{\phi}\left(\frac{1}{z}-1\right)\log(1-\phi z)\right].$$

- which gives us the useful relation

$$\sum_{k=0}^{\infty}\frac{\partial^i p(k;m,\phi)}{\partial m^i}z^k = \left(\sum_{k=0}^{\infty}h_k z^k\right)^i\left(\sum_{k=0}^{\infty}p(k;m,\phi)z^k\right)$$

with

$$\left.\begin{aligned} h_0 &= -1 \\ h_k &= \phi^{k-1}\left(\frac{1}{k}-\frac{\phi}{k+1}\right) \quad (k\geq 1)\end{aligned}\right\}.$$

¶Now Newton-Raphson could be implemented

- a feasible algorithm for derivatives

$$\left.\begin{aligned} p^{(1)}(k;m,\phi) &= h_k * p(k;m,\phi) \\ p^{(2)}(k;m,\phi) &= h_k * p^{(1)}(k;m,\phi) \end{aligned}\right\}.$$

- a statistician's old friend

$$\left.\begin{aligned} U(m,\phi;X) &= \frac{\partial l}{\partial m} = \sum_{i=1}^{n} \frac{p^{(1)}(X_i;m,\phi)}{p(X_i;m,\phi)} \\ J(m,\phi;X) &= -\frac{\partial^2 l}{\partial m^2} = \sum_{i=1}^{n} \left[\left(\frac{p^{(1)}(X_i;m,\phi)}{p(X_i;m,\phi)} \right)^2 - \frac{p^{(2)}(X_i;m,\phi)}{p(X_i;m,\phi)} \right] \end{aligned}\right\}.$$

- what an easy job to do point and interval estimation, e.g.

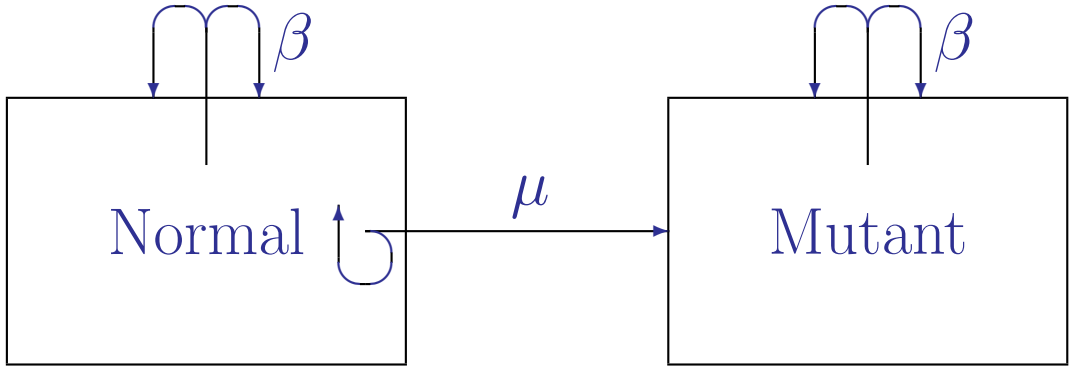
$$\tilde{m}_{k+1} = \tilde{m}_k + \frac{U(\tilde{m}_k,\phi;X)}{J(\tilde{m}_k,\phi;X)}.$$

¶This simple idea can be reused, many times

- Bartlett derived another p.g.f., for a completely-random model

$$G(z;\alpha,\phi) = \left[\frac{(1-\phi)z}{1-\phi z-(1-z)(1-\phi z)^{\alpha}}\right]^{N_0}$$

∞ • α, the mutation rate can be similarly estimated



¶asymptotically, the two models are equivalent (2007)

- if $X \sim LD(m, \phi)$, then

$$Pr(X = n) \sim \frac{\phi^n}{\Gamma\left(\frac{1-\phi}{\phi} m\right)} \frac{1}{n^{1-m(1-\phi)/\phi}}.$$

- if $Y \sim B(\alpha, \phi)$ with $N_0 = k$ initial nonmutant cells

$$Pr(Y = n) \sim \frac{\phi^n}{\Gamma(k\alpha)} \frac{1}{n^{1-k\alpha}}.$$

- if $m = \frac{\alpha\phi}{1-\phi} N_0$, then

$$\lim_{n \to \infty} \frac{Pr(X = n)}{Pr(Y = n)} = 1$$

¶With Haldane's model, we don't even have a p.g.f.

- To find $p(g;k)$, Haldane suggested finding g integers, $a_0, a_1, \ldots, a_{g-1}$, such that

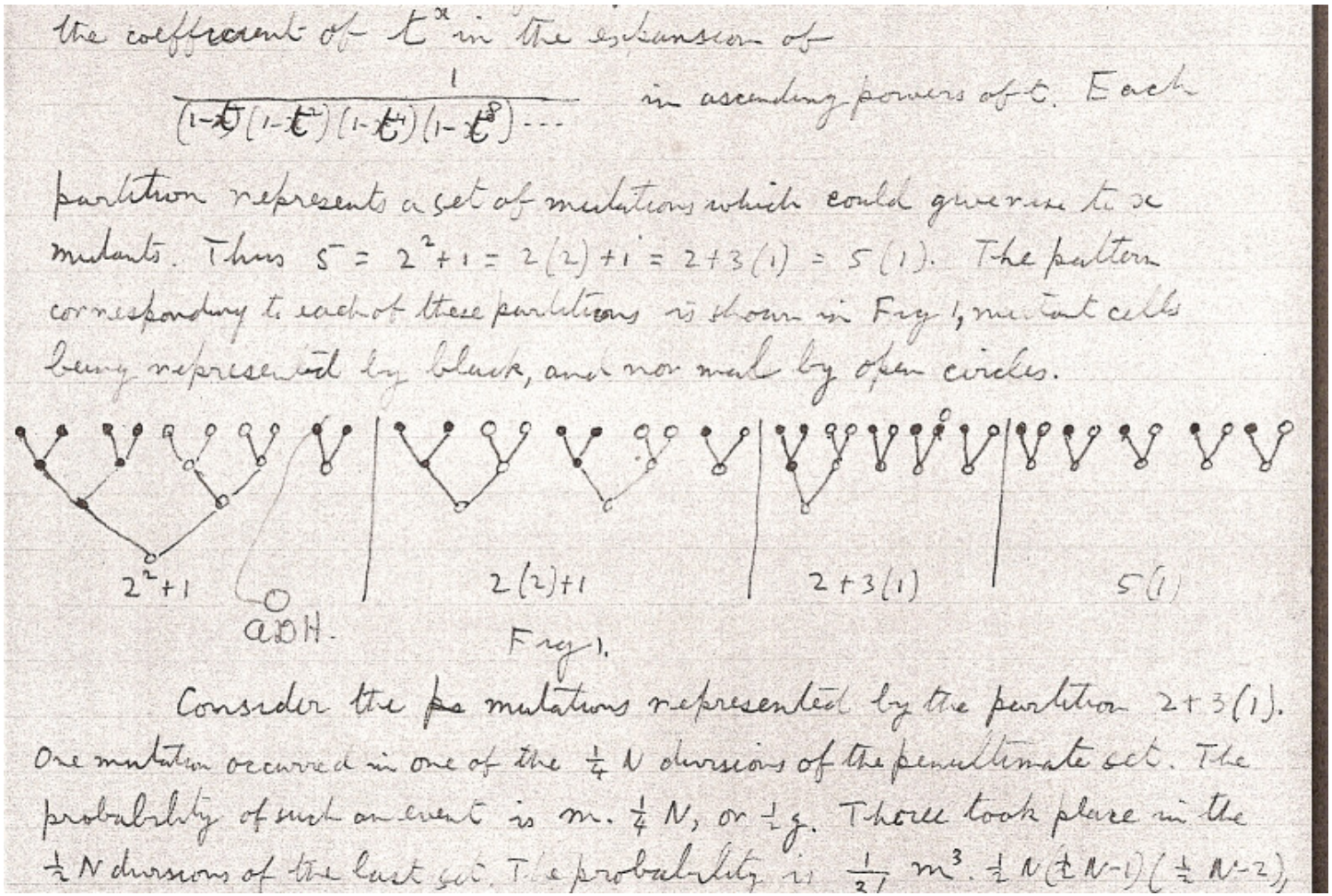
$$k = a_0 2^0 + a_1 2^1 + \ldots + a_{g-1} 2^{g-1}$$

- actually, some constraints must be imposed.

$$a_i \le 2^{g-1-i} - \sum_{k=1}^{g-1-i} 2^{g-1-i-k} a_{g-k} \stackrel{\text{def}}{=} a_i^* \quad (i = 0, \ldots, g-1)$$

- example: if $g = 6$ and $k = 45$, we have 89134 partitions to consider; only 524 of them satisfy the first condition, and only 374 satisfy both conditions.

¶Haldane's manuscript was unearthed in 1991



the coefficient of t^x in the expansion of

$$\frac{1}{(1-t)(1-t^2)(1-t^4)(1-t^8)\cdots}$$

in ascending powers of t. Each partition represents a set of mutations which could give rise to 2c mutants. Thus $5^- = 2^2+1 = 2(2)+1 = 2+3(1) = 5(1)$. The pattern corresponding to each of these partitions is shown in Fig 1, mutant cells being represented by blank, and non mut by open circles.

2^2+1 aBH. $2(2)+1$ $2+3(1)$ $5(1)$

Fig 1.

Consider the 2a mutations represented by the partition $2+3(1)$. One mutation occurred in one of the $\frac{1}{4}N$ divisions of the penultimate set. The probability of such an event is $m \cdot \frac{1}{4}N$, or $\frac{1}{2}g$. Three took place in the $\frac{1}{2}N$ divisions of the last set. The probability is $\frac{1}{2!}m^3 \cdot \frac{1}{2}N(\frac{1}{2}N-1)(\frac{1}{2}N-2)$,

* currently archived by University College London

¶The same idea works even when we don't have a p.g.f. (2006)

- from the Markovian property of the process, we have

$$p(g+1;k) = \sum_{j=\max(0,k-N_g)}^{\lfloor k/2 \rfloor} P(Y_{g+1}=k|Y_g=j)p(g;j)$$

$$= \sum_{j=\max(0,k-N_g)}^{\lfloor k/2 \rfloor} \binom{N_g-j}{k-2j} \mu^{k-2j}(1-\mu)^{N_g-k+j} p(g;j)$$

- this simplifies computation of the probability mass function
- this allows derivatives to be computed
- this allows the implementation of the Newton-Raphson

¶an example

- in Demerec's experiment: $N_0 = 90$ and $N_T = 1.9 \times 10^8$.
- thus, $\phi = 1 - 90/(1.9 \times 10^8)$ and $g \approx 21$.
- data from *Proc. Natl. Acad. Sci. USA* 31:16-24 (1945).

33	18	839	47	13	126	48	80	9	71
196	66	28	17	27	37	126	33	12	44
28	67	730	168	44	50	583	23	17	24

- Lea & Coulson: $\hat{\mu}_\beta = 5.71 \times 10^{-8}$ and an asymptotic 95% CI, $(4.55 \times 10^{-8}, 6.94 \times 10^{-8})$
- Bartlett: $\hat{\alpha} = 5.78 \times 10^{-8}$ CI=$(4.58 \times 10^{-8}, 7.07 \times 10^{-8})$
- Haldane: $\hat{\mu} = 7.14 \times 10^{-8}$ CI=$(5.94 \times 10^{-8}, 8.39 \times 10^{-8})$

¶Latest developments

- if X is thinned by a "thinning" probability ε, the distribution is

$$G_Y(z) = \exp\left(m\xi\frac{(1-z)\log[\varepsilon(1-z)]}{1+\xi z}\right)$$

where

$$\xi = \frac{\varepsilon}{1-\varepsilon}$$

- if we take an appropriate limiting process, the distribution is

$$G(z; A, k) = \left(\frac{1}{1 - A(z^{-1}-1)\log(1-z)}\right)^k$$

¶Does there exist another formulation?

J.F. Crow, Genetics 124:207-211 (1990)

> Taking advantage of my newly formed acquaintance with Fisher, I asked him how to find the distribution of mutant cells ... He leaned back in his chair, thought for perphas a minute, and wrote a generating function ... I took the paper ... and then managed to lose it.