

PBG 200A Notes

Sam Fleischer

October 31, 2016

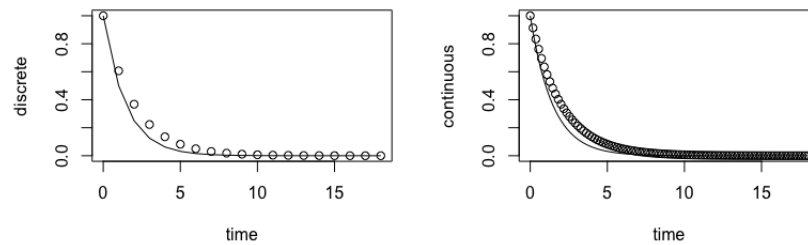
1 Questions from the Notes

Question 1

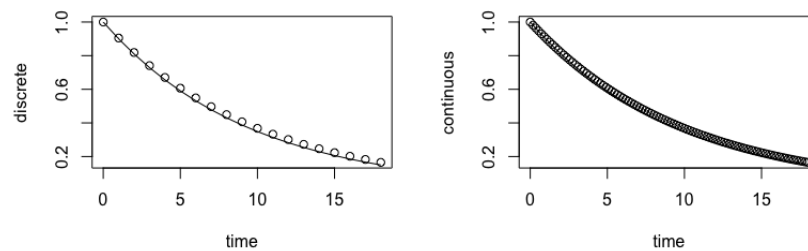
$$H_t = H_0 \exp\left(-\frac{t}{2N}\right)$$
$$0.0049 = 0.005 \exp\left(-\frac{200/3}{2N}\right)$$
$$\ln 0.98 = -\frac{200}{6N}$$
$$N = -\frac{200}{6 \ln 0.98}$$
$$N = 1650$$

Question 2

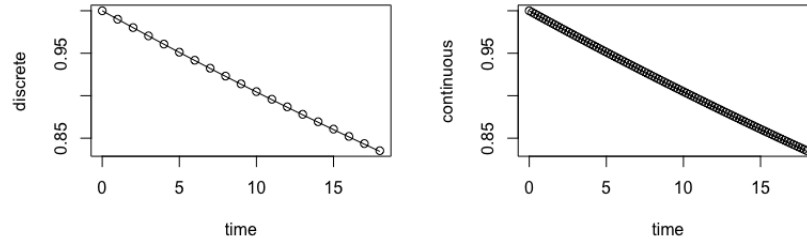
$x = 0.5$:



$x = 0.1$:



$x = 0.01$:



Question 3

- A) $\pi_1 = 0.0012$, $\pi_2 = 0.000014$?????
- B) $N_i = \frac{\theta_i}{4\mu}$ for $i = 1, 2$. If $\mu = 2 \times 10^{-8}$ and $\theta_1 = 0.0012$, then $N_1 = 15,000$. If $\theta_2 = 0.000014$, then $N_2 = 175$.
- C) Oh my God they're adorable.

2 In-Class Notes

2.1 Review

- Neutral polymorphism/alleles
- Only 2% encodes for proteins
- changes outside exons may be neutral if they do not disrupt regulatory sites
- examples of potentially neutral alleles:
 - synonymous changes in codons
 - non-synonymous changes which replaces one amino acid with a functionally similar one
 - non-synonymous change which produces a large change in a phenotype on which selection no longer acts

2.2 Loss of Heterozygosity

- genetic drifts happens at a rate inversely proportional to population size (no genetic drift in infinite population size, i.e. Hardy-Weinberg)
-

$$H_t = \left(1 - \frac{1}{2N}\right) \times H_{t-1} + \frac{1}{2N} \times 0$$

$$H_t = \left(1 - \frac{1}{2N}\right)^t H_0$$

If $\frac{1}{2N} \ll 1$, then $1 - \frac{1}{2N} \approx \exp\left(-\frac{1}{2N}\right)$, and thus $H_t \approx \exp\left(-\frac{t}{2N}\right)H_0$.

2.2.1 Examples

- the chief of Pingelap island was heterozygous for colorblindness
- Ashkenazi Jews and Tay-Sachs disease
- extra fingers in Amish population

2.3 Mutation can maintain nonzero equilibrium heterozygosity

We can write down the probability of no mutation between a pair of sequences,

$$\begin{aligned}
\mathbb{P}[\text{no mutation between sequences}] &= \frac{1}{2N}(1-\mu)^2 + \frac{1}{2N}\left(1-\frac{1}{2N}\right)(1-\mu)^4 + \frac{1}{2N}\left(1-\frac{1}{2N}\right)^2(1-\mu)^6 \\
&= \frac{1}{2N} \sum_{t=0}^{\infty} \left[\left(1-\frac{1}{2N}\right)^t (1-\mu)^{2t+2} \right] \\
&\approx \frac{1}{2N} \sum_{t=0}^{\infty} \exp\left[-\frac{t}{2N} - 2t\mu\right] \\
&\approx \frac{1}{2N} \int_0^{\infty} \exp\left[-\frac{t}{2N} - 2t\mu\right]
\end{aligned}$$

where μ is the rate of mutation and N are the number of alleles. This has a nice answer:

$$\mathbb{P}[\text{no mutation between sequences}] = \frac{1}{2N} \left(\frac{1}{\frac{1}{2N} + 2\mu} \right) = \frac{1}{1 + 4N\mu}$$

Given that $4N\mu \ll 1$, we have $H = \frac{4N\mu}{1 + 4N\mu} \approx 4N\mu = \theta$. The rate of genetic drift is slower in larger populations, so there is more of a chance for mutations to take hold.

2.4 Effective population size

The size of an ideal population N_e in which drift occurs at the same rate as that in an actual population N :

$$N_e \ll N$$

Bottlenecks so greatly affect the rate of loss of heterozygosity that the “effective” population size is close to the bottleneck sizes, not the large population size. Harmonic mean is very sensitive to small numbers.

2.5 Coalescence

We have

$$t \sim \text{Geom}\left(\frac{1}{2N}\right), \quad \implies \quad \mathbb{E}[t] = 2N \text{ generations} \quad (1)$$

Often we assume $t \sim \text{Exp}\left(\frac{1}{2N}\right)$. Given a mutation rate of μ , then the number of mutations from two lineages going back t generations is $2\mu t$, so

$$\pi := \mathbb{E}[\text{pairwise differences}] = 2\mu \mathbb{E}[t] = 2\mu 2N = 4N\mu = \theta \quad (2)$$

π is the observation - θ is the theoretical value.