Probabilistic approach to the probability of improvement in Fisher's geometric model

Yoav Ram

Hadany Lab
December 2013

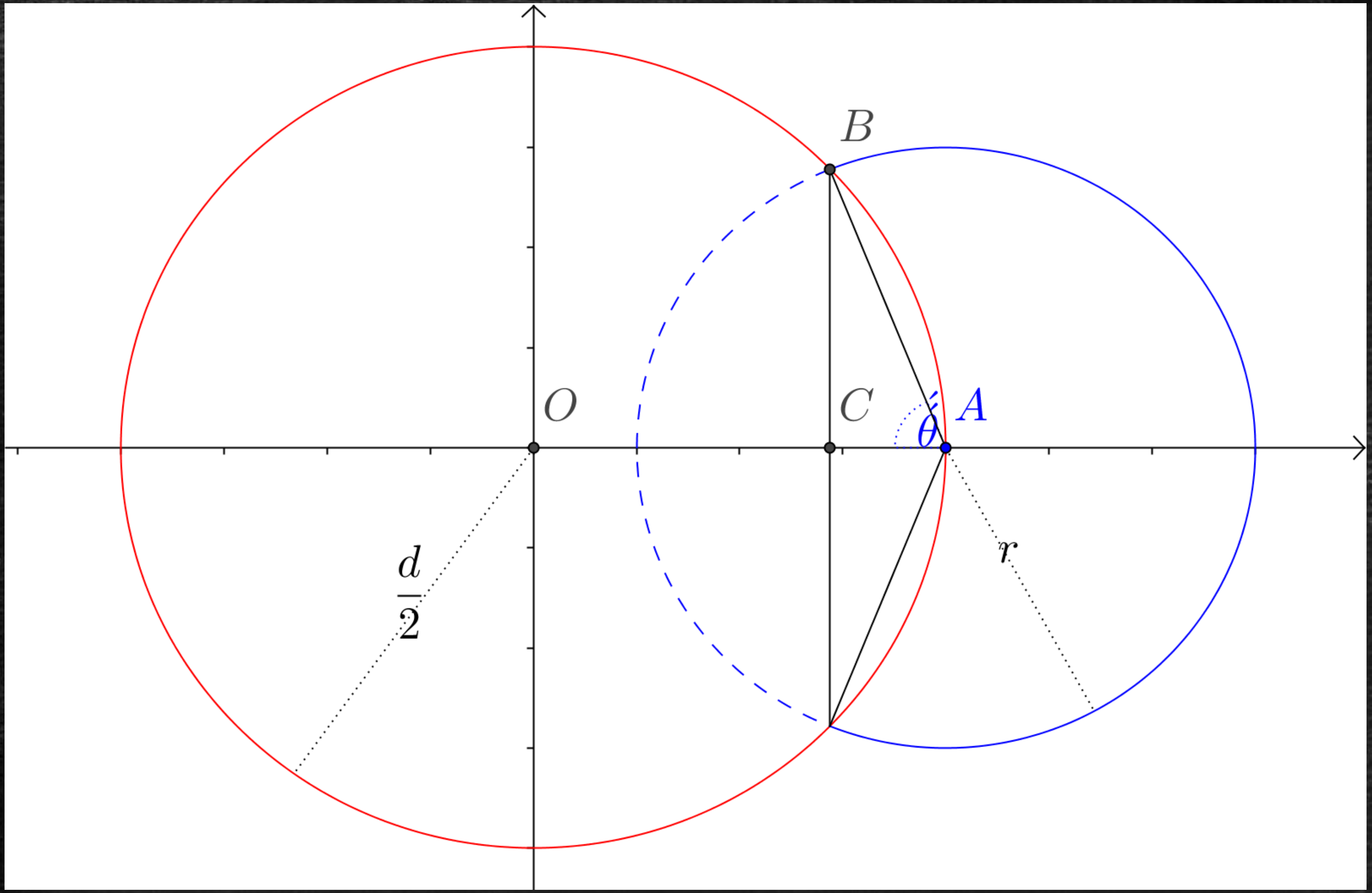
Introduction

- Fisher introduced his "geometric model" in page 32 of the book *The Genetical Theory of Natural Selection*
- Download pdf from http://archive.org/details/geneticaltheory031631mbp
- **Mutation pleiotropy**
- Support for **adaptation by small mutations**
- Criticisms:
  - Wright: adaptive peak shifts
  - Kimura: fixation probability

Sir Ronald A. Fisher
1890-1962
England & Australia

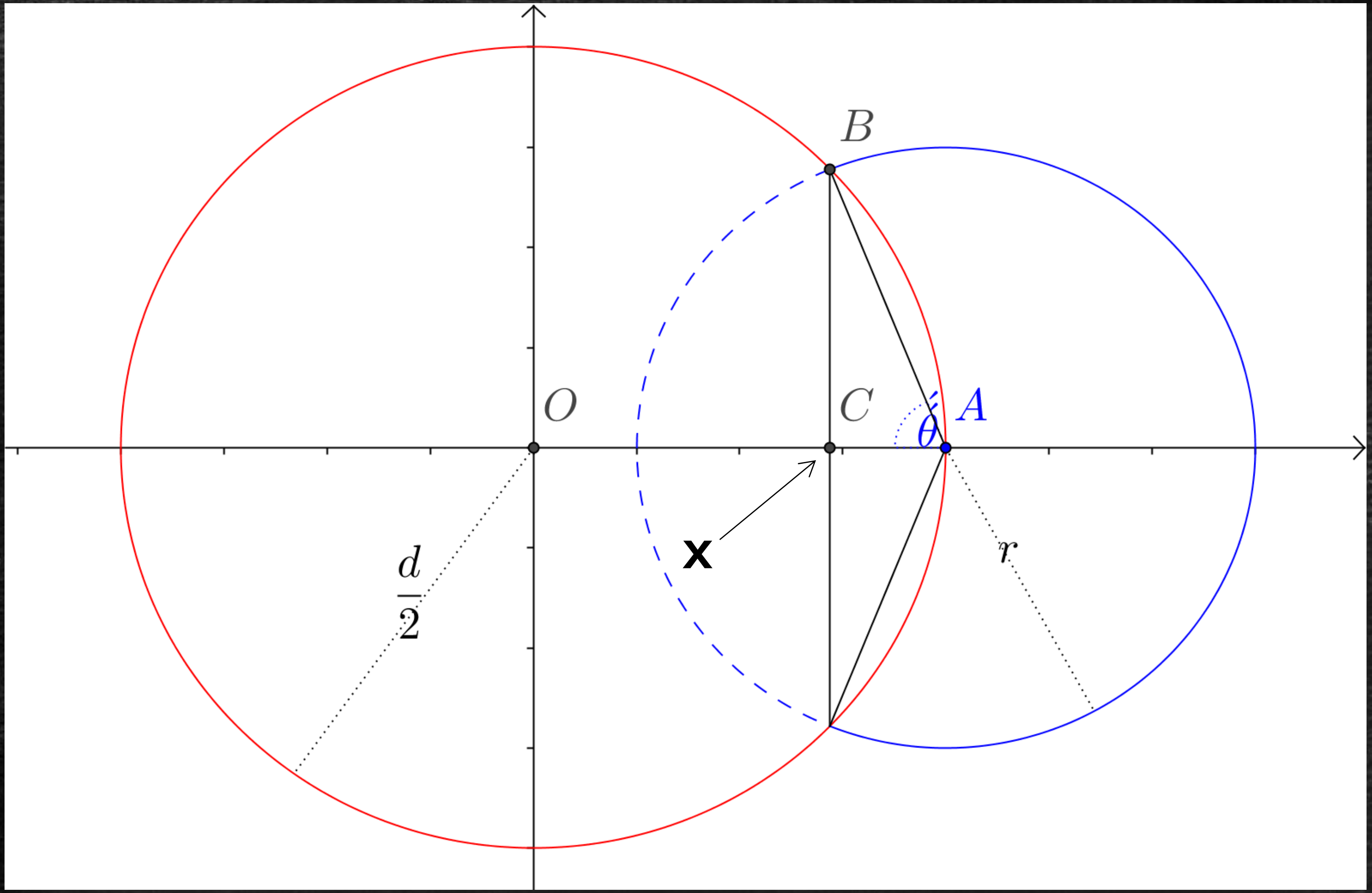
Fisher's geometric model in 2D

- 2 trait phenotype space
- *O*: optimal phenotype
- *A*: current phenotype
- Blue circle: all possible mutants
- Red circle: all phenotypes with fitness = A
- Dashed arc: mutants with fitness > A



Probability of improvement, p_{2D}

- p_{2D} is the part of the blue circle that is dashed
- B: intersection of the circles
- $C = (x, 0)$
- θ is the angle between AB and AC
- $p_{2D} = \frac{2\theta}{2\pi}$



Probability of improvement, p_{2D}

We need to find *x:*

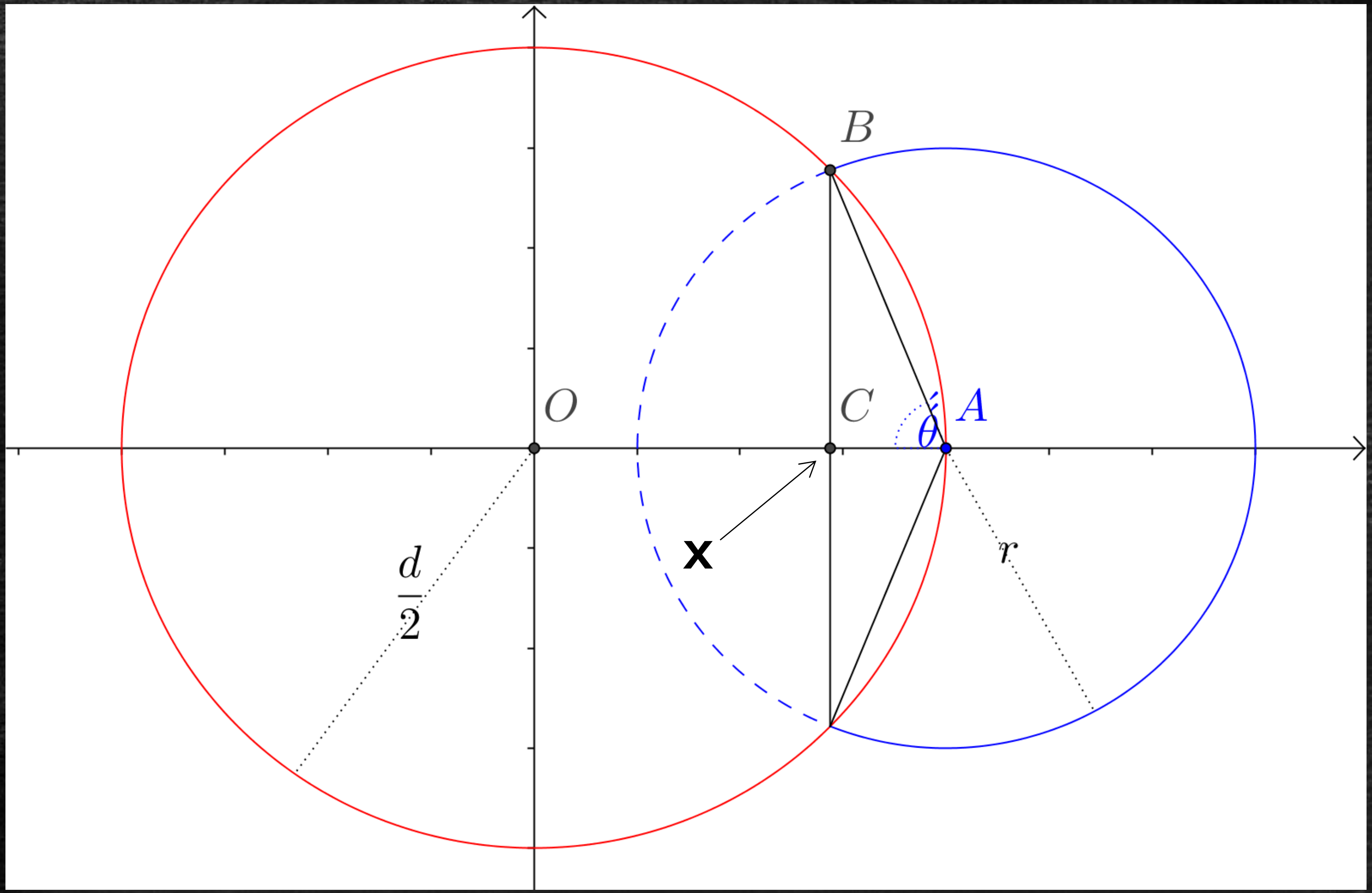
$$\begin{cases} \left(x - \frac{d}{2}\right)^2 + y^2 = r^2 \\ x^2 + y^2 = \left(\frac{d}{2}\right)^2 \end{cases}$$

$$x = \frac{d}{2} - \frac{r^2}{d}$$

$$AB = r$$

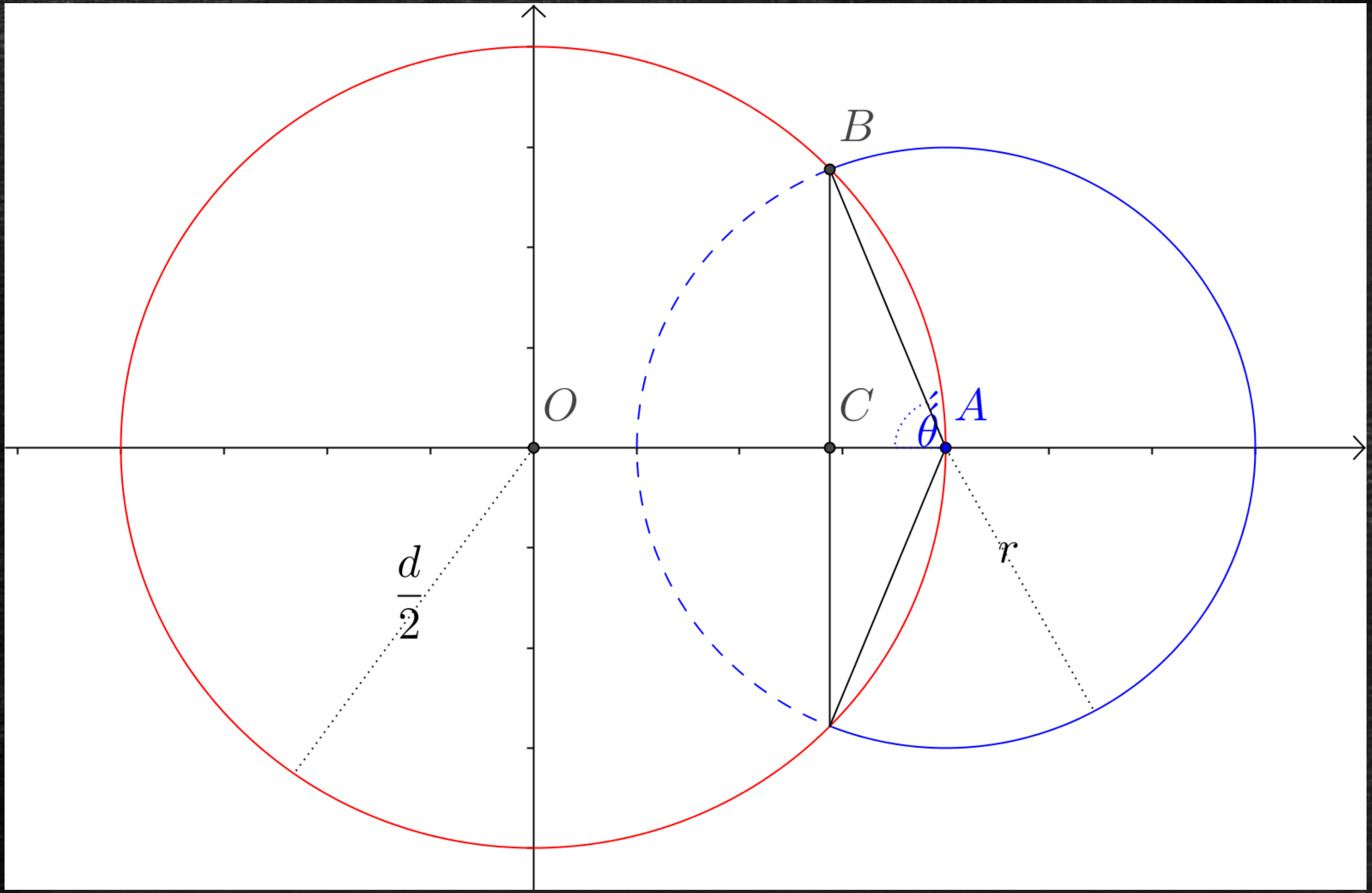
$$AC = d/2 - x$$

$$\cos\theta = \frac{AC}{AB} = \frac{r}{d}$$



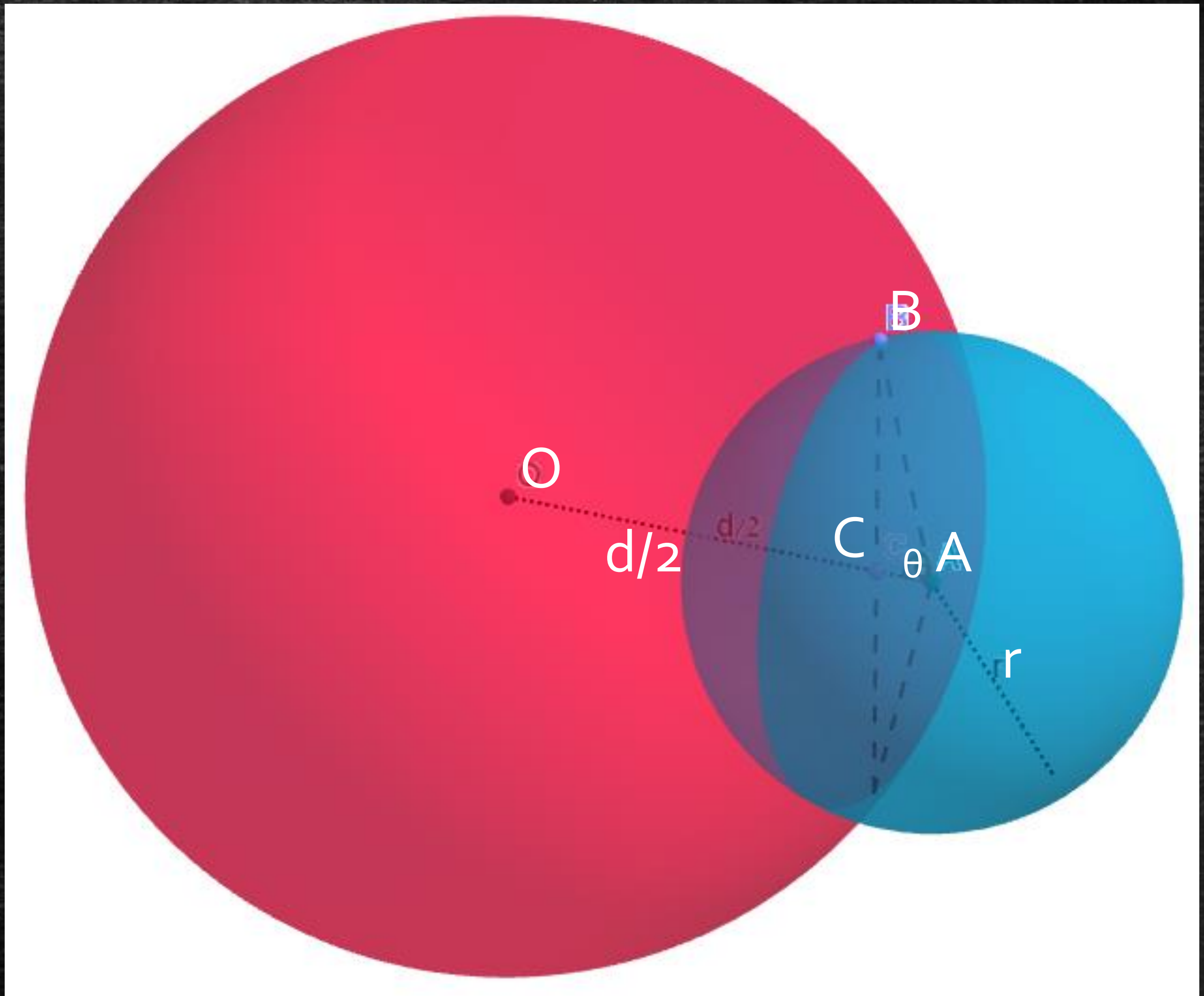
Probability of improvement, p_{2D}

$$p_{2D} = \frac{\cos^{-1}\left(\frac{r}{d}\right)}{\pi}$$



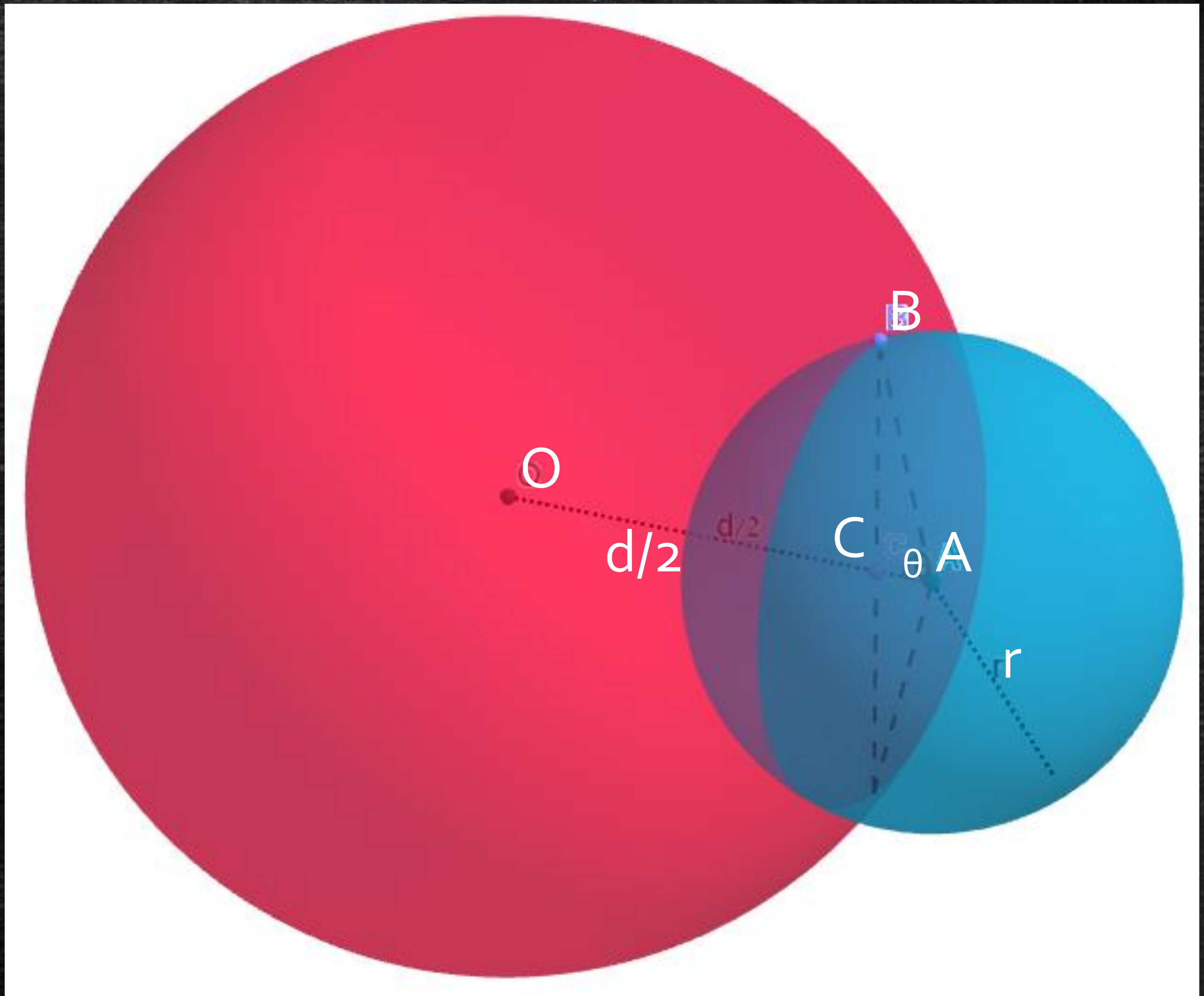
Fisher's geometric model in 3D

- 3 trait phenotype space
- Find AC like in 2D
- The area of a **spherical cap** – the shaded part of the blue sphere – is $2\pi rh$ - where h is r-AC
- The area of the whole sphere is $4\pi r^2$
- $p_{3D} = \frac{2\pi r\left(r-\frac{r^2}{d}\right)}{4\pi r^2}$



Fisher's geometric model in 3D

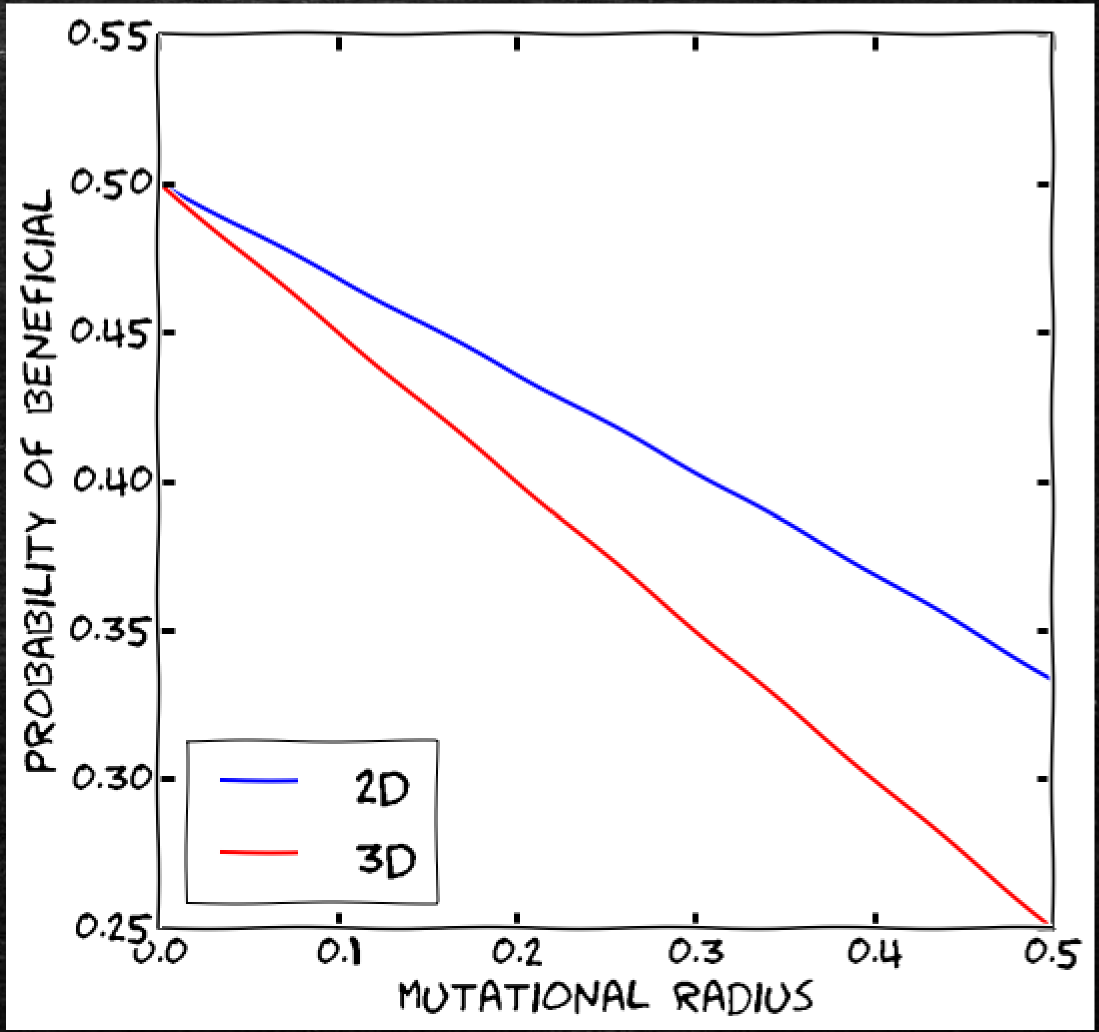
$$p_{3D} = \frac{1}{2}\left(1 - \frac{r^2}{d}\right)$$



Comparison: 2D vs. 3D

$$p_{2D} = \frac{\cos^{-1}\left(\frac{r}{d}\right)}{\pi}$$

$$p_{3D} = \frac{1}{2}\left(1 - \frac{r^2}{d}\right)$$



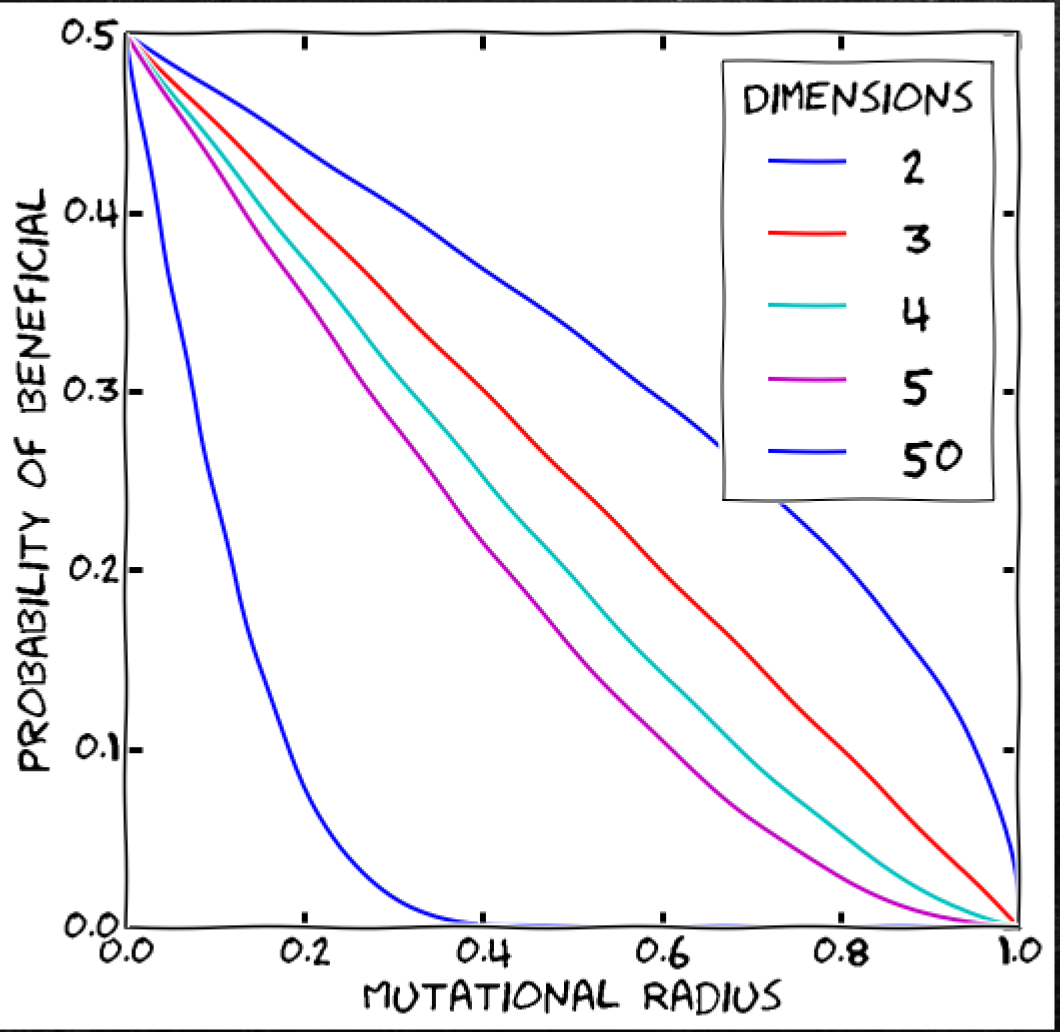
General result

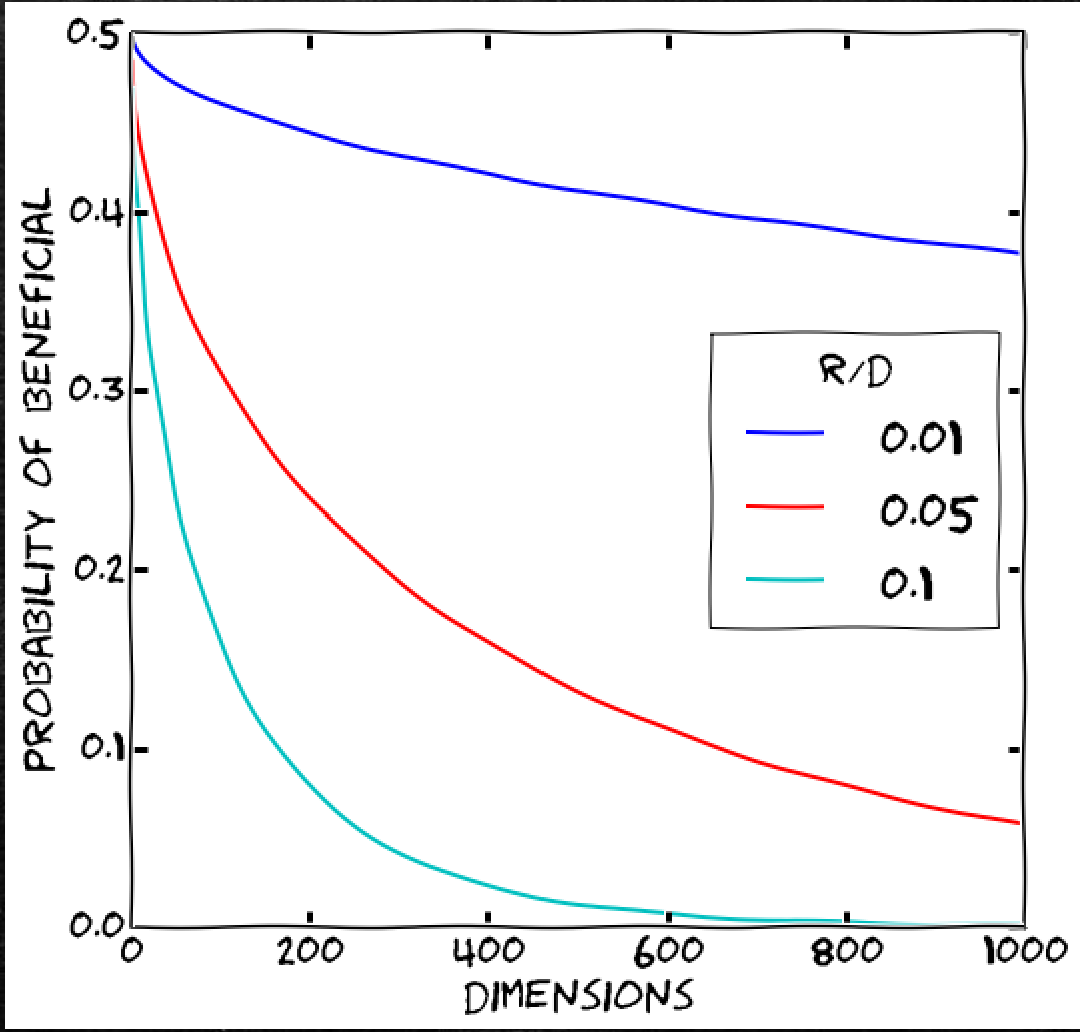
- For an arbitrary number of traits *n*

$$p_n = \frac{\int_0^{\cos^{-1}\left(\frac{r}{d}\right)} \sin^{n-2}(\theta) d\theta}{\int_0^{\pi} \sin^{n-2}(\theta) d\theta}.$$

Rice 1990, Hartl & Taubes 1996, Waxman & Welch 2005

Multiple traits





Asymptotic result for many traits

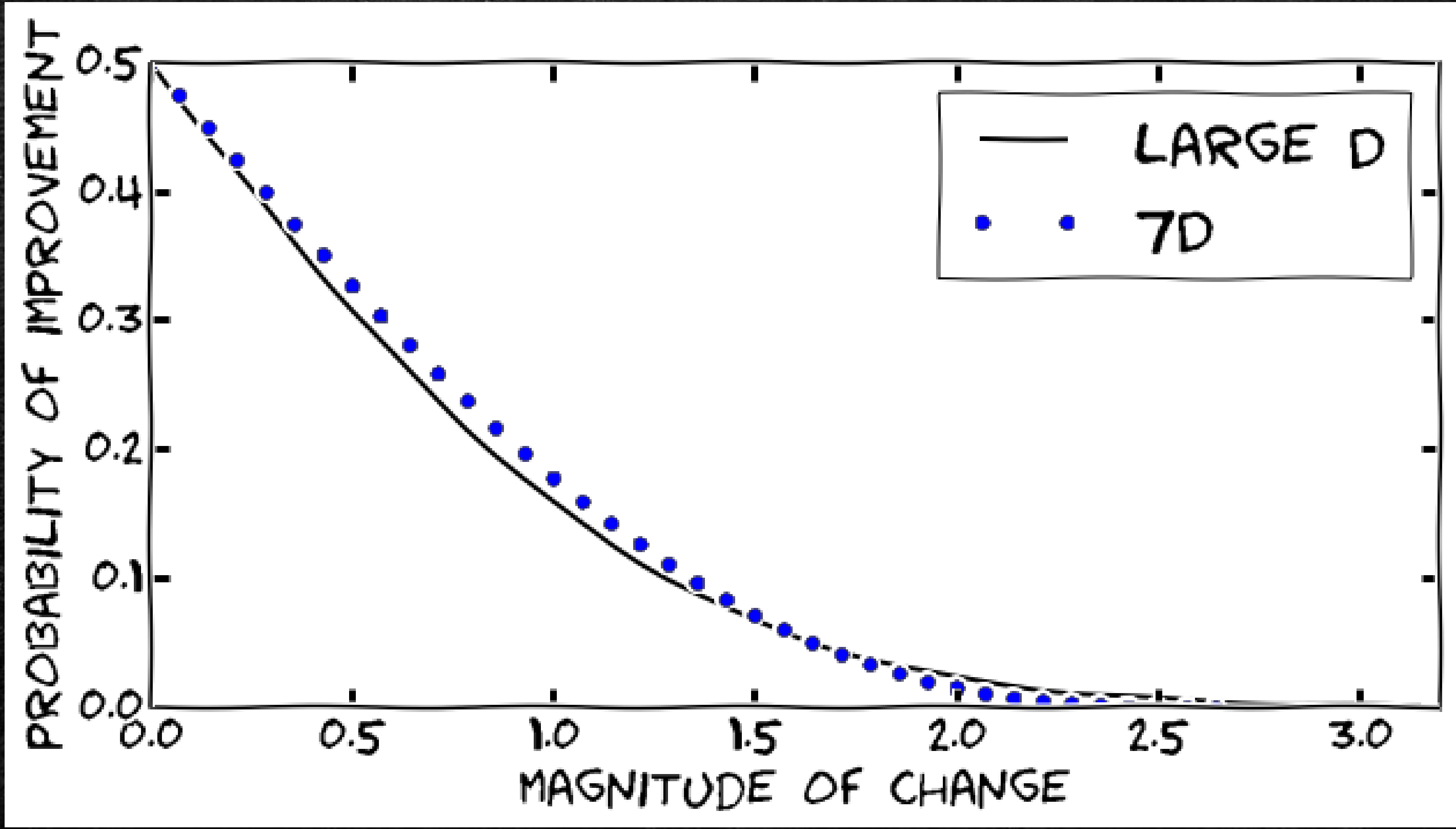
- This is the result given by Fisher (1930, pg. 40)
- Can be found as an approximation of p_n

$$p_F = 1 - \phi\left(\sqrt{n}\frac{r}{d}\right) =$$

$$\frac{1}{\sqrt{2\pi}}\int_{\sqrt{n}\frac{r}{d}}^{\infty} e^{-\frac{t^2}{2}} dt$$

ϕ is the cumulative probability function of the standard normal distribution

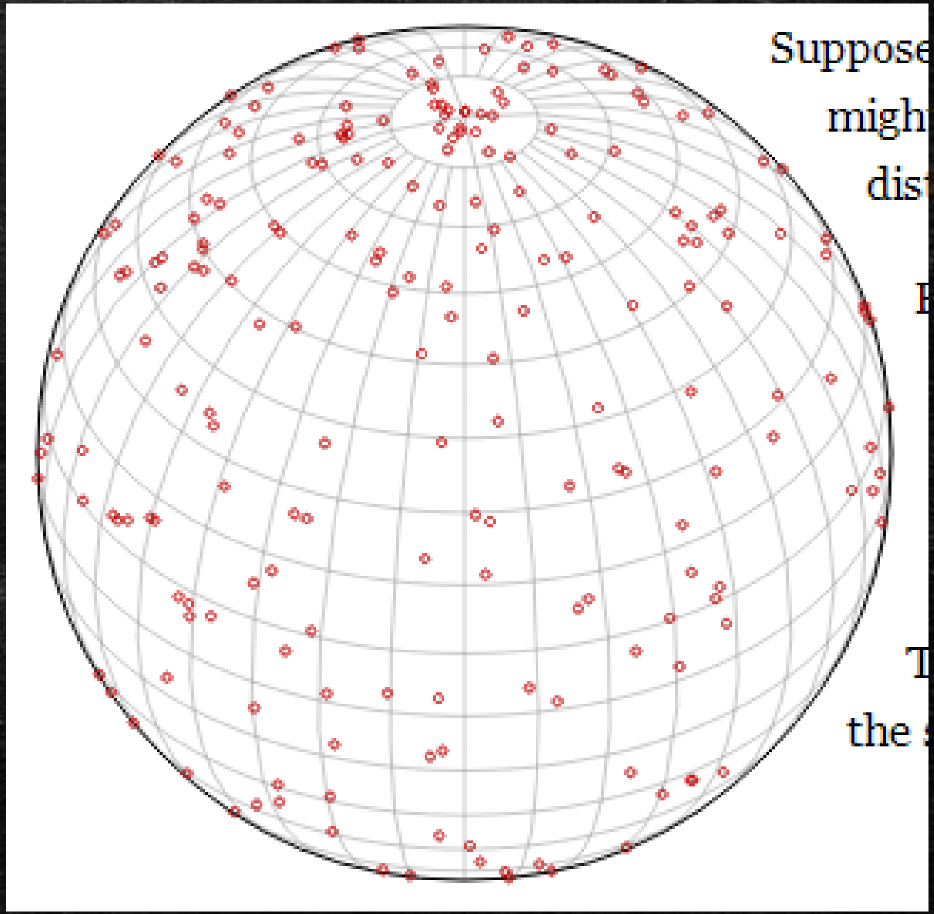
Comparison of asymptotic result



Random points on a sphere

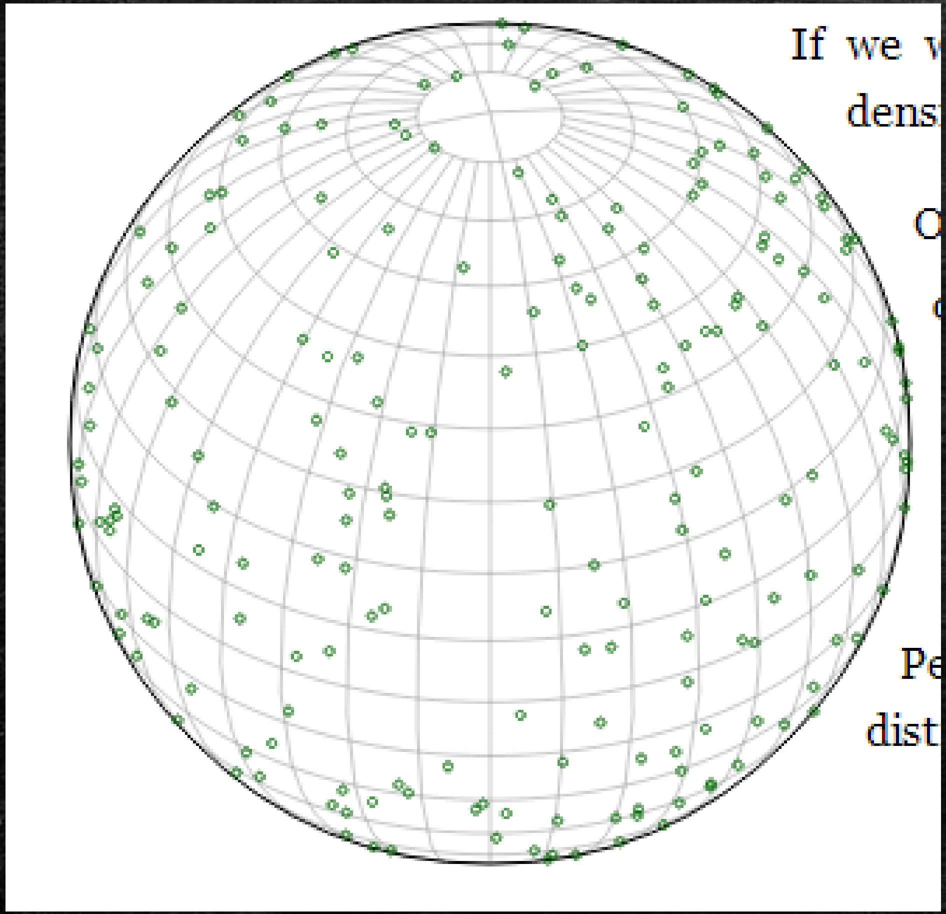
- Generate uniformly distributed points on a sphere.
- Picking **spherical coordinates**:
  - $\lambda \sim Uniform\,[-180°, 180°)$
  - $\varphi \sim Uniform\,[-90°, 90°)$
- **Uneven distribution**: density increasing around the poles

- The area of a given “square” of width $\Delta\lambda$ and height$\Delta\varphi$ varies with φ
- See how the squares get smaller towards the poles?



http://www.jasondavies.com/maps/random-points/

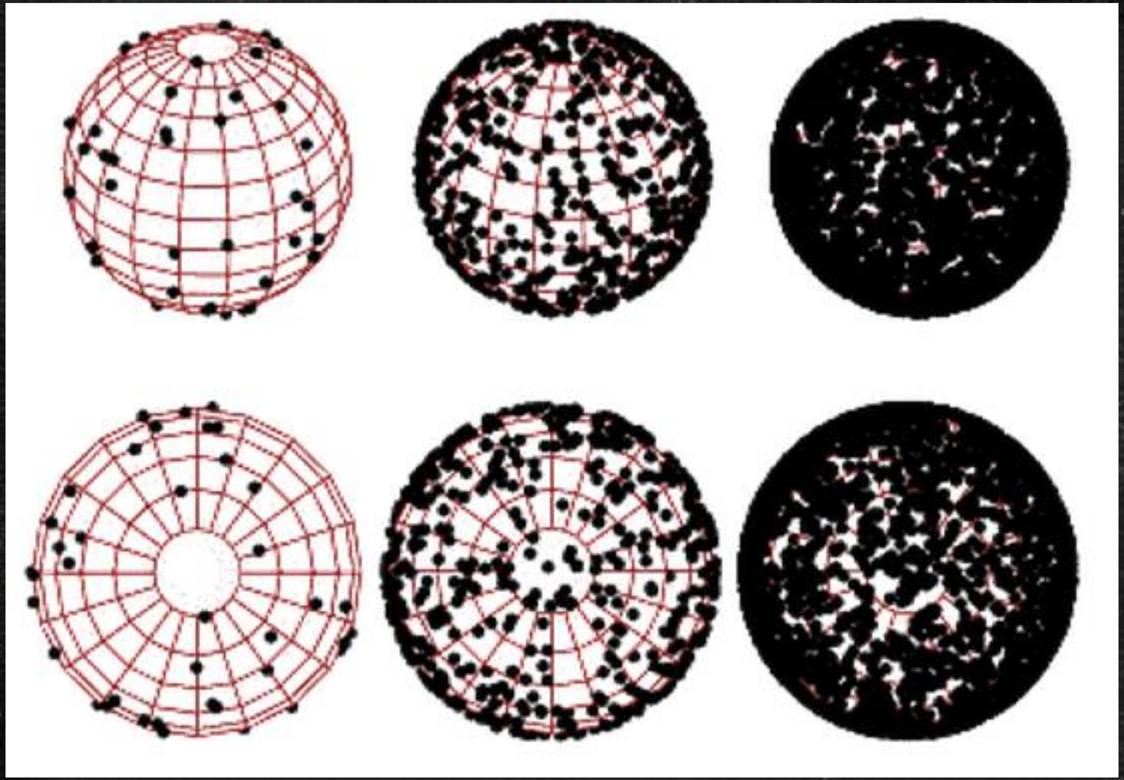
Random points on a sphere



- Any area on the sphere should contain approximately the same density of points
- $\lambda \sim Uniform\ [-180°, 180°)$
- $x \sim Uniform\ [0, 1)$
- $\varphi\ = \cos^{-1}(2x\ -\ 1)$
- Some points seem too close together, and some seem too far apart.

Random points on a hypersphere

- For a hypersphere in *n* dimensional space and radius *r*:
- Pick $Z_i \sim Normal(0{,}1),\ \ 1 \leq i \leq n$
- Normalize and strech: $X_i = r \cdot \frac{Z_i}{\|Z\|} = r \cdot \frac{Z_i}{\sum_{i=1}^{n} Z_i^2}$
- X is a random point on the hypersphere with radius 1



http://mathworld.wolfram.com/SpherePointPicking.html

Muller, 1959

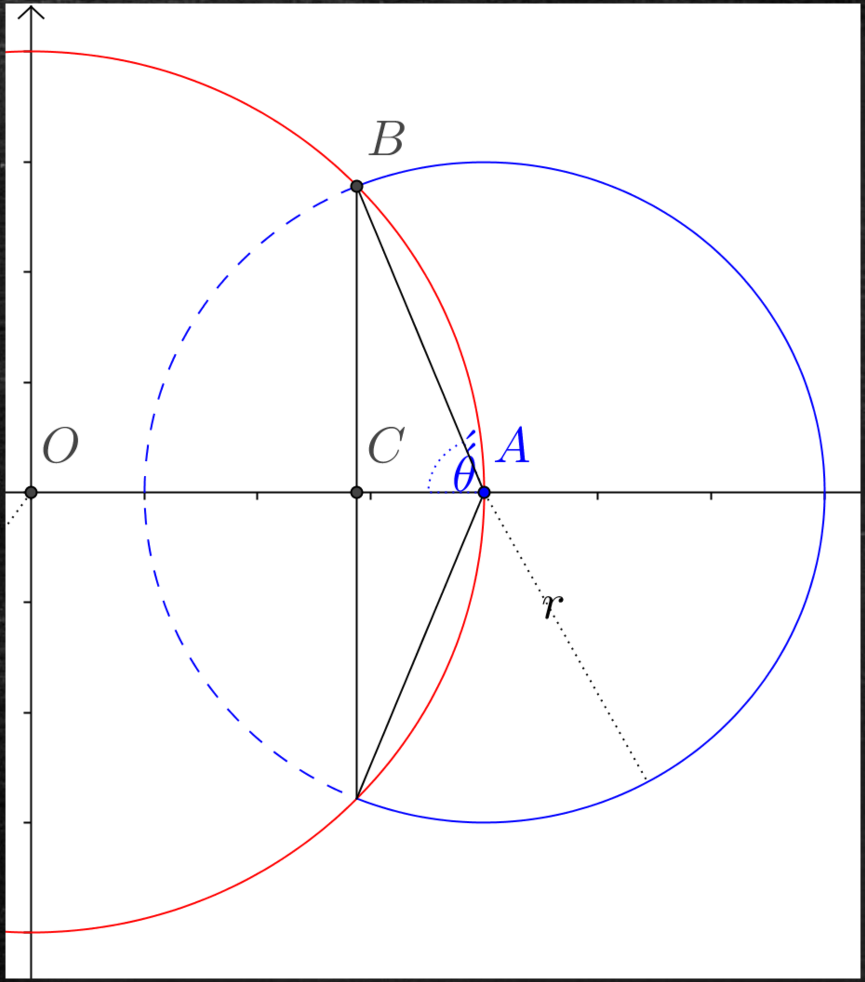
Our probabilistic approach

$A = \left(\frac{d}{2}, 0, \dots, 0\right)$ – current phenotype

$X = r \cdot \frac{Z}{\|Z\|}$ - mutation

Probability of improvement:

$$p = \Pr\left(\|A + X\| < \frac{d}{2}\right)$$
$$= Pr\left(\frac{X_1}{r} > \frac{r}{d}\right)$$



Relation to previous results

$$Pr\left(\frac{X_1}{r} > \frac{r}{d}\right) = \frac{1}{2} Pr\left(\frac{X_1^2}{r^2} > \left(\frac{r}{d}\right)^2\right) = \frac{1}{2}\mathrm{Pr}\left(\frac{\mathrm{Z}_1^2}{\mathrm{Z}_1^2 + \sum_{i=2}^{n} Z_i^2} > \frac{\mathrm{r}^2}{\mathrm{d}^2}\right)$$

The sum of squares of normals is chi-squared:

$$\sum_{i=1}^{k} Z_i^2 \sim \chi^2(k)$$

The ratio of chi-squared and its sum with another chi-squared is beta:

$$\frac{\mathrm{Z}_1^2}{\mathrm{Z}_1^2 + \sum_{i=2}^{n} Z_i^2} \sim Beta\left(\frac{1}{2}, \frac{n-1}{2}\right)$$

The probability of improvement is:

$$Pr\left(\frac{X_1}{r} > \frac{r}{d}\right) = \frac{1}{2}\left(1 - I_{\frac{r^2}{d^2}}\left(\frac{1}{2}, \frac{n-1}{2}\right)\right)$$

Relation to previous results

Using the hypergeometric function ${}_2F_1$:

$$\int_0^{\phi} \sin^n(\theta)\, d\theta = \frac{1}{2} B_{\sin^2(\phi)}\left(\frac{n+1}{2}, \frac{1}{2}\right)$$
$$= \frac{1}{2} I_{\sin^2(\phi)}\left(\frac{n+1}{2}, \frac{1}{2}\right) B\left(\frac{n+1}{2}, \frac{1}{2}\right)$$

Therefore we can substitute to get:

$$p_n = \frac{1}{2}\left(1 - I_{\frac{r^2}{d^2}}\left(\frac{1}{2}, \frac{n-1}{2}\right)\right) = Pr\left(\frac{X_1}{r} > \frac{r}{d}\right)$$

Relation to previous results

We use the identity $I_x(a,a) = \frac{1}{2} I_{4x(1-x)}\left(a, \frac{1}{2}\right), 0 \le x \le \frac{1}{2}$:

$$p_n = \frac{1}{2}\left(1 - I_{\frac{r^2}{d^2}}\left(\frac{1}{2}, \frac{n-1}{2}\right)\right) = I_{\frac{1}{2}\left(1-\frac{r}{d}\right)}\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$$

$$= P\left(D < \frac{1}{2}\left(1 - \frac{r}{d}\right)\right), \qquad \mathrm{D} \sim Beta\left(\frac{n-1}{2}, \frac{n-1}{2}\right)$$

Generally, $Beta(a,a)$ is well approximated by $Normal\left(\frac{1}{2}, \frac{1}{8a+4}\right)$ when a is large

Relation to previous results

D is approximated by $Normal\left(\frac{1}{2}, \frac{1}{4n}\right)$

and

$$p_n \approx \phi\left(\frac{\frac{1}{2}\left(1-\frac{r}{d}\right)-\frac{1}{2}}{\sqrt{\frac{1}{4n}}}\right) = \phi\left(-\sqrt{n}\frac{r}{d}\right)$$

$$= 1 - \phi\left(\sqrt{n}\frac{r}{d}\right)$$

This is the result presented by Fisher

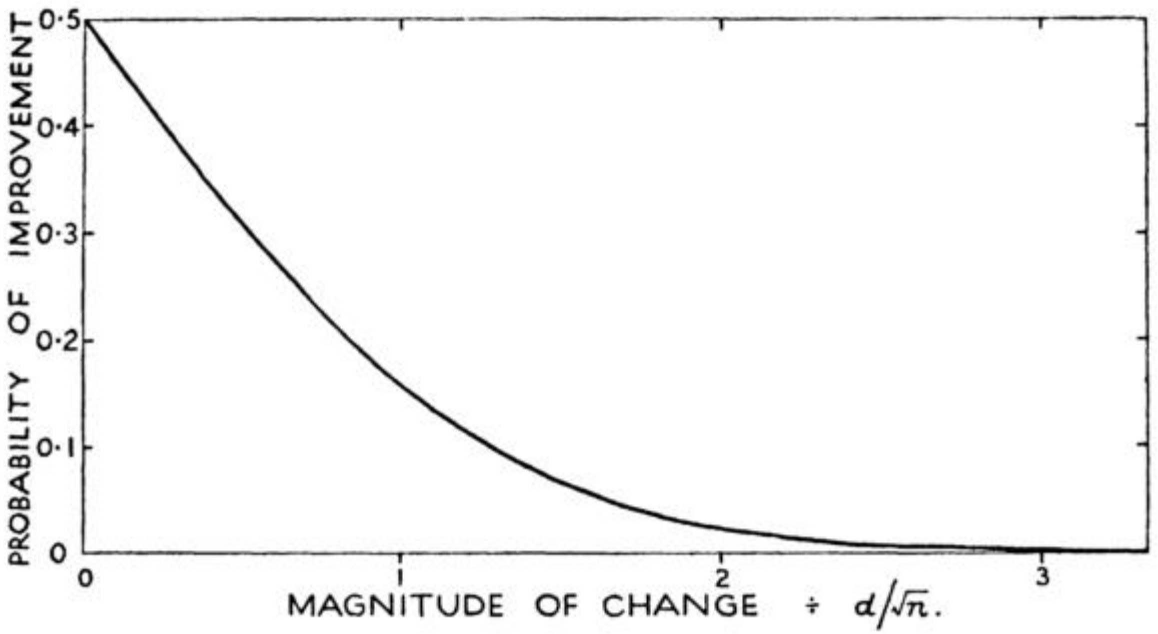


FIG. 3. The relation between the magnitude of an undirected change and the probability of improving adaptation, where the number of dimensions (n) is large

$$p = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-\frac{1}{2}t^2}dt,\ x = r\sqrt{n}/d.$$

New intuition

$$P_n = Pr\left(\frac{X_1}{r} > \frac{r}{d}\right)$$

For a mutation to improve fitness, the **relative size of the effect of the mutation** in the direction towards the optimum must be **larger** than half the **ratio** between the total **mutation size** and the **distance to the optimum**