

# 未検出アレルの頻度推定 と 観察ジェノタイプの珍しさ

法数学勉強会

2018/01/20

京都大学 統計遺伝学分野 山田 亮

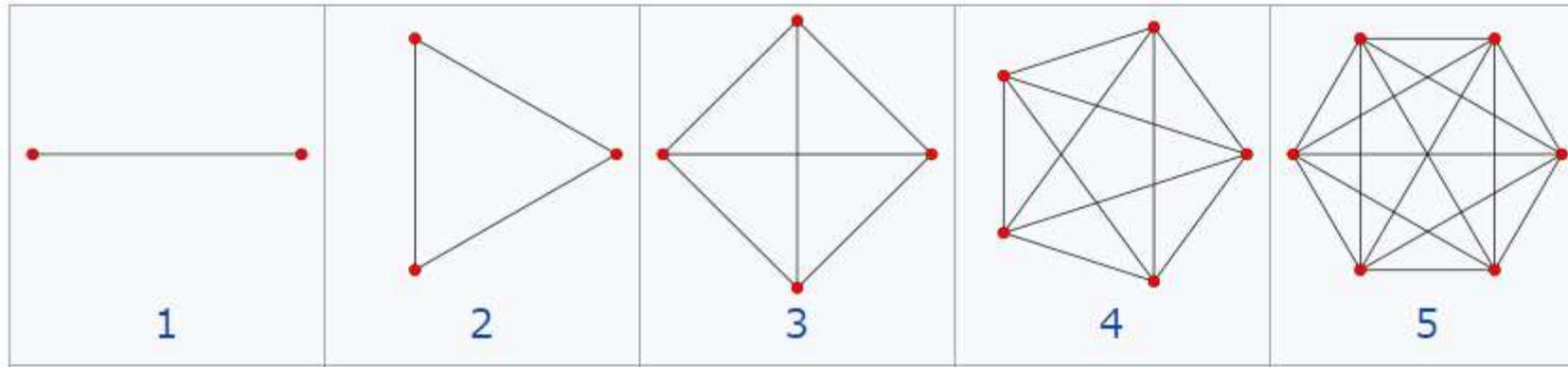
# ディリクレ分布・ディリクレ過程

- 前回の話

# 頻度推定をすること

- 観察するのは整数
- 推定するのは実数

# 頻度推定と、三角形、四面体、単体



# 2つのアレルの観測数

## 2つのアレルの頻度推定

- $P_1^{n_1} P_2^{n_2} = P_1^{n_1} (1-P_1)^{n_2}$
- 確率
- 尤度
- 微分する：最尤推定

# 手を動かそう

- P1を横軸にする
- P2はどこか
  
- P1を横軸に、P2を縦軸にする
- $P1 + P2 = 1$  : どこか

3つのアレルの観測数

3つのアレルのアレル頻度推定

- $P_1^{n_1} P_2^{n_2} P_3^{n_3}$
- 確率
- 尤度
- 微分する：最尤推定

# 手を動かそう

- P1をx軸に、P2をy軸に、P3をz軸に取る
- $P1 + P2 + P3 = 1$  : どこか
- P1, P2, P3 を平面に取る

# 4つのアレルの観測数

## 4つのアレルの頻度推定

- $P_1^{n_1} P_2^{n_2} P_3^{n_3} P_4^{n_4}$
- 確率
- 尤度
- 微分する：最尤推定

# 手を動かそう

- P1をx軸に、P2をy軸に、P3をz軸に、P4をw軸に取る
- $P1 + P2 + P3 + P4 = 1$  : どこか
- P1, P2, P3, P4 を 3次元空間に取る

3つのアレル ( $A_1, A_2, A_3$ )

プールすれば ( $B_1 = A_1, B_2 = A_2 + A_3$ )

- プールするってどういうこと？

絵で理解しよう

4つのアレル(A1,A2,A3,A4)

プールすれば( $B1=A1, B2=A2+A3+A4$ )

# 観測数、ゼロ

- 観測数、ゼロのアレル頻度はゼロか？
- 2 アレル
  - 観測数、 $(0,0)$ のとき
  - 観測数、 $(1,0)$ のとき
  - 観測数、 $(4,0)$ のとき
  - 観測数、 $(1000,0)$ のとき

未観測アレルは、あと、いくつある？

- SNPの場合
- STRの場合
- Y染色体ハプロタイプの場合

# 観測数、 $(3,2)$ のとき

- SNPの場合
- STRの場合
- Y染色体ハプロタイプの場合

# 観察ジェノタイプの珍しさ

- 2アレル  $P1, P2$  のとき
- $n1, n2$  と観測する確率とその珍しさ
  
- 珍しければ、 $P1, P2$  という仮定が間違っていた？
- 仮説の棄却

# 観察ジェノタイプの珍しさ

- 2アレルの過去の観察データが  $m_1, m_2$  であったとき
- $n_1, n_2$  と観測する確率とその珍しさ
- $P_1, P_2$  を推定しよう。その上で、 $n_1, n_2$  と観察する確率を計算しよう。珍しさも数値にしよう

# 観察ジェノタイプの珍しさ

- 2アレルの過去の観察データが  $m_1, m_2$  であったとき
- $n_1, n_2$  と観測する確率とその珍しさ
- $m_1, m_2$  の観察と、 $n_1, n_2$  の観察が、同じ母集団からなのかどうか、も検定できる
  - 漸近近似検定、正確確率検定

観察ジェノタイプの珍しさ  
多アレルの場合

存在の知られていなかったアレルが含まれているとき

- SNP、集団
- SNP、個人のタイピング、次世代シーケンシング

# 存在の知られていなかったアレルが含まれているとき

- 「変なデータなのではないか？」
  - どうして「変」だと思うのか？
  - どれくらい「変」なのか？
- 
- 「変」でも、他に説明がなければ、認めることになる
  - 「対案」は何か？
  - 「変」対「変ではなくなる対案」：尤度比