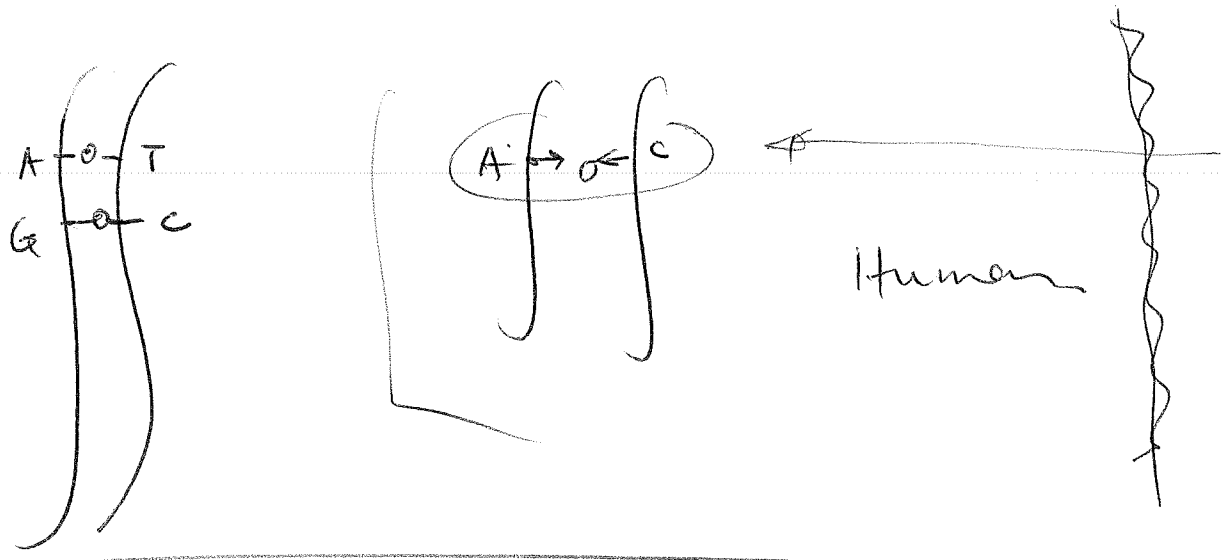


diploids

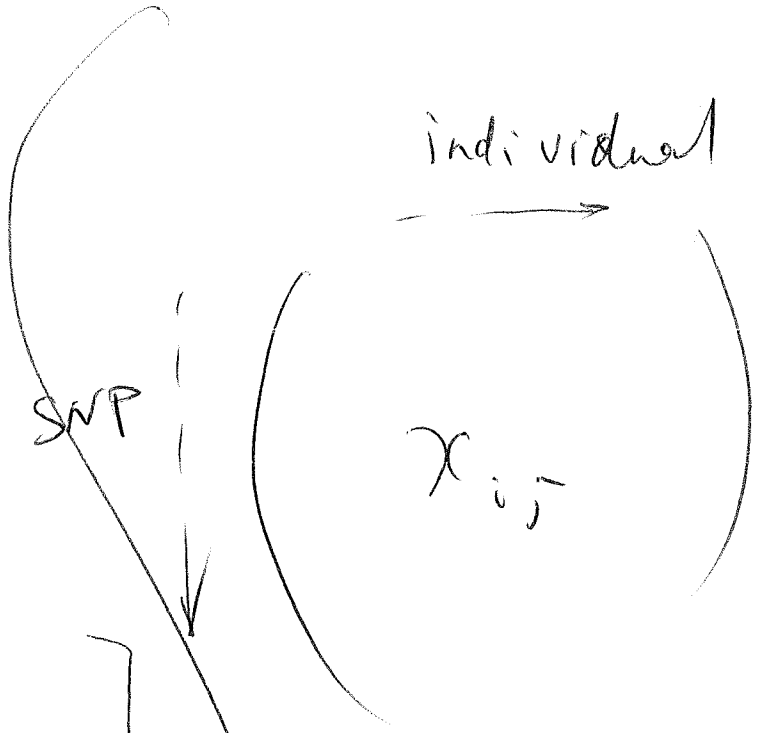
P_1 P_2

P_1 ⇢ a

P_2 → A

Descendant

a a	→ 0
a A	→ 1
A A	→ 2



$i = 1, \ldots, m$

$m = 3 \times 10^6$

$j = 1, \ldots, n$ (12,500)

$n \approx 12,500$

individual 2/03

SNP ~10^6

$$\left(x_{ij} \right)$$

y_j < 1 disease, 0 not

$$\Pr(y_j = 1 \mid x_{1j}, x_{2j}, \ldots, x_{mj}) = p_j$$

$$\log\left(\frac{p_j}{1-p_j}\right) = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_m x_{mj}$$

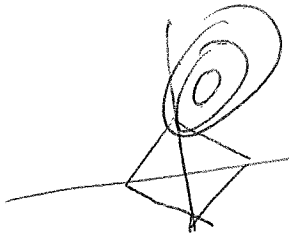
~~only a few~~ $\beta_0, \beta_1, \ldots, \beta_m$

only a few of are $\neq 0$

$\{\beta_1, \ldots, \beta_{s_0}\}$ none 0

$\beta_{s_0+1}, \ldots, \beta_m$ all 0

lasso

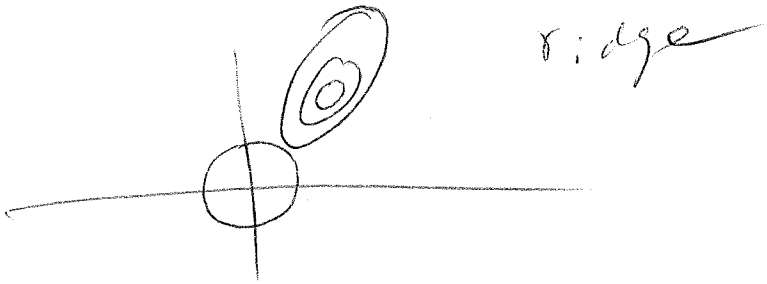


indi - -

snp (x_{ij}) y_i < 1 / 0

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1j} + \cdots + \beta_m x_{mj}$$

SNP1 SNP2

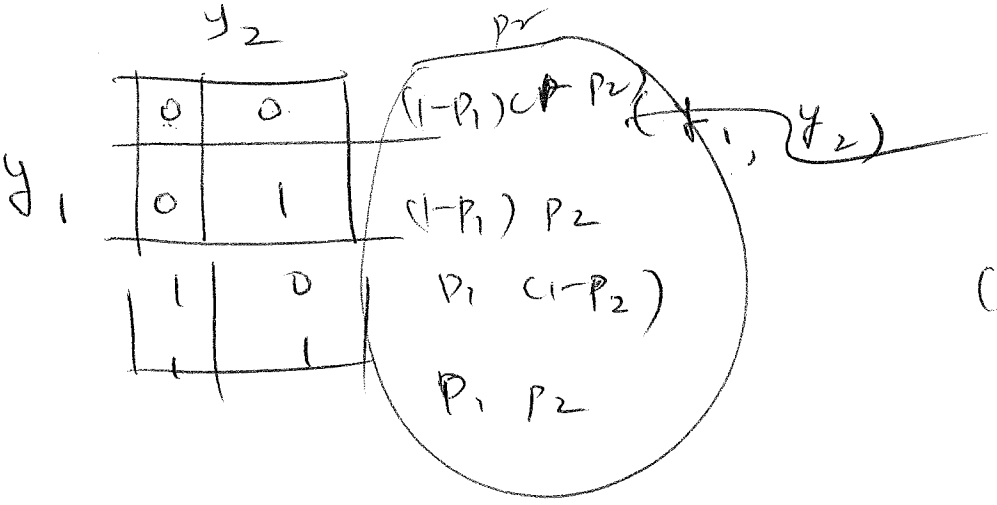


$\beta_1, \cdots, \beta_{100}$ $\beta_{101}, \cdots \beta_{200}$ ⟶

GWAS

$y_1 \sim f_1 = p_1^{y_1} (1-p_1)^{1-y_1}$

$y_2 \sim f_2 = p_2^{y_2} (1-p_2)^{1-y_2}$



$f_1 \times f_2 = \underbrace{p_1^{y_1} (1-p_1)^{1-y_1}}\ \underbrace{p_2^{y_2} (1-p_2)^{1-y_2}}$

$$= \begin{cases} (1-p_1)(1-p_2) \\ (1-p_1)\ p_2 \\ p_1\ (1-p_2) \\ p_1\ p_2 \end{cases}$$

$y_i < \begin{matrix} 1 & \text{Cancer} \\ 0 & \text{not} \end{matrix}$

$\Pr(y_i = 1) = p_i$ depends on $x_{1i}, x_{2i}, \ldots, x_{mi}$

$\Pr(y_i = 0) = 1 - p_i$

$$f_i = p_i^{y_i} (1 - p_i)^{1 - y_i}$$

$$= \begin{cases} p_i & y_i = 1 \\ 1 - p_i & y_i = 0 \end{cases}$$

Φ $y_1, \ldots, y_n$ are ind. $\Leftrightarrow$

$$g(y_1, \ldots, y_n) = \prod_{i=1}^{n} f_i$$

Via induction.