Week 5 Variant Calling

Monday, 14 February 2022   8:01 AM

unbiased coin toss

$P(H) = \frac{1}{2}$ $P(T) = \frac{1}{2}$

$$P(H) + P(T) = 1$$

Bias = $P(H)$

$$\{0, 0.2, 0.5, 0.8, 1\}$$

Bayes theorem

Likelihood

PRIORS

$$\underbrace{P(A|B)}_{\text{posterior}} = \frac{\overbrace{P(B|A)}^{\text{Likelihood}} P(A)}{P(B)}$$

$$P(A|B) \propto P(B|A)\, P(A)$$

having observed H

1/5

$$P(Bias|H) = \frac{P(H|Bias) \times P(Bias)}{P(H)}$$

if bias = 0 $\Rightarrow$ $P(H|bias) = 0$

$$P(H) = P(H|bias = 0) \times P(\overset{=0}{bias})$$

$$+ P(H \mid bias = 0.2) \times P(bias = 0.2)$$

$$+ \dots + P(H \mid bias = 1) \times P(bias = 1)$$

$$P(bias = 0.2 \mid H) = \frac{P(H \mid bias = 0.2) \times P(bias)}{P(H)}$$

$$= \frac{0.2 \times 0.2}{0 \times \frac{1}{5} + 0.2 \times \frac{1}{5} + 0.5 \times \frac{1}{5} + 0.8 \times \frac{1}{5} + 1 \times \frac{1}{5}}$$

$$= 0.08$$

$$P(bias = 0.5 \mid H) = \frac{0.5 \times 0.2}{}$$

$$= 0.2$$

$$P(bias = 0.2 \mid HH) = \frac{\overset{0.2}{P(HH \mid bias)} \times \overset{0.08}{\underline{P(bias)}}}{P(HH)}$$

$$= 0.032$$

Reference genome $\rightsquigarrow$ A allele [TRUTH]

Other than A $\rightarrow$ ERROR

n = # of reads

k = # of reference alleles

$\Rightarrow$ $n-k$ = # of errors

BERNOULLI TRIAL $\qquad$ $P = P(H)$

P(n experiments ; k of something)

$$= {}^{n}C_{k}\, p^{k} (1-p)^{n-k}$$

$$= \frac{n!}{k!\,(n-k)!}$$ P^k (1-p)^n-k

eg. $n = 10$; $k = 2$

$$P(n=10; k=2) = {}^{10}C_{2}\,(0.5)^2 (0.5)^8$$

$P(AA \mid \text{reads}) = ? \propto P(\text{reads} \mid AA) \times P(AA)$

$P(AB \mid \text{reads}) \propto P(\text{reads} \mid AB) \times P(AB)$

$P(BB \mid \text{reads}) \propto P(\text{reads} \mid BB) \times P(BB)$

one locus $\rightarrow$ $\{A, C\}$ biallelic

$n =$ individuals ; $2n =$ alleles

k A alleles ; $2n - k$ C allele

$$P(A) = \frac{k}{2n} = p \quad ; \quad P(C) = 1 - P(A) = \frac{2n-k}{2n} = q$$

	$p(A) = p$ A	$p(C) = q$ C
$p(A) = p$ A	$P(AA) = p^2$	$P(AC) = pq$
$p(C) = q$ C	$P(AC) = pq$	$P(CC) = q^2$

$$\left.\begin{aligned} P(AA) &= p^2 \\ P(CC) &= q^2 \\ P(AC) &= pq + pq = 2pq \end{aligned}\right]$$

eg. $n = 50$; $2n = 100$

$k = 20$, $n - k = 80$

$$P(A) = p = \frac{20}{100} = 0.2 \qquad q = 1 - p = 0.8$$

$$\Rightarrow P(AA) = p^2 = 0.2^2 = 0.04$$

$$P(AC) = 2pq = 2 \times 0.2 \times 0.8 = 0.32$$

$$P(CC) = q^2 = 0.8^2 = 0.64$$

$$P(\text{reads} \mid AA) = {}^{n}C_{k}\, \epsilon^{k} (1-\epsilon)^{n-k}$$

ϵ = prob. of 'error'
(not reference allele)

$$P(\text{reads} \mid BB) = {}^{n}C_{n-k}\, \epsilon^{n-k} (1-\epsilon)^{k}$$

$$P(\text{reads} \mid AB) = \,? = {}^{n}C_{k} \left(\frac{1}{2^{n}}\right)$$

$n = 30 \qquad k = 3 \qquad \Rightarrow n-k = 27$

$P(AA \mid n=30, k=3) = ?$ | $P(AA) = 0.25$
$P(AB \mid n=30, k=3) = ?$ | $P(AB) = 0.5$
$P(BB \mid n=30, k=3) = ?$ | $P(BB) = 0.25$

$$P(AA \mid n=30, k=27) \propto P(\text{reads} \mid AA) \times P(AA)$$

$$\propto {}^{30}C_{27} \left(\frac{3}{30}\right)^{27} \left(\frac{27}{30}\right)^{30-27} \times 0.25$$

$$\propto 7.99 \times 10^{-25}$$

$\times\ 0.5$

$$P(AB|\text{reads}) \propto C_k \left(\frac{1}{2^n}\right) \times$$

$$\propto {}^{30}C_{27} \left(\frac{1}{2^{30}}\right) \times 0.5$$

$$\propto 1.89 \times 10^{-6}$$

$$P(BB|\text{reads}) = {}^{n}C_k \, \epsilon^k (1-\epsilon)^{n-k}$$

$$= {}^{30}C_3 \, (0.9)^3 \, (0.1)^{27} \times 0.25$$

$$= \underline{0.059}$$

$$\binom{n}{k}$$