

# NOTAS DE CLASE

## MARCADORES GENÉTICOS MOLECULARES (FIT-616)

*Dr. Humberto Reyes Valdés*

### Medidas de polimorfismo genético

---

#### PIC

Sean X, Y y Z tres individuos. X es el individuo de referencia, el cual se aparea con Y. Z es una progenie individual. La probabilidad de que en una población de apareamiento aleatorio sea posible distinguir el haplotipo heredado de X a Z en un locus marcador codominante, se denomina PIC "polymorphism information content" (contenido de información por polimorfismo).

Esto sucede cuando ocurren cualquiera de las dos siguientes situaciones:

X es heterocigoto y Y tiene un genotipo diferente a X

X es heterocigoto, Y es un heterocigoto igual a X, y Z es homocigoto

Por lo tanto, 1-PIC es la probabilidad de que ninguna de las situaciones anteriores ocurra. Es decir, cuando (1) X es homocigoto, o cuando (2) X y Y son heterocigotos iguales y Z es heterocigoto. Estos dos eventos son mutuamente excluyentes y la suma de sus probabilidades es 1-PIC.

Sea  $p_i$  la frecuencia del alelo  $i$ , y  $a$  el número de alelos en la población

Para el evento (1)

$$P1 = \text{Sum}[p_i^2, \{i, 1, a\}]$$

$$\sum_{i=1}^a p_i^2$$

Para el evento (2)

$$P2 = \text{Sum}[(1/2) * (2 p_i p_j)^2, \{i, 1, a-1\}, \{j, i+1, a\}]$$

$$\sum_{i=1}^{a-1} \sum_{j=i+1}^a \frac{1}{2} (2 p_i p_j)^2$$

Y el PIC es

$$pic = 1 - P1 - P2$$

$$-\sum_{i=1}^a p_i^2 - \sum_{i=1}^{a-1} \sum_{j=i+1}^a \frac{1}{2} (2 p_i p_j)^2 + 1$$

$$PIC = 1 - \sum_{i=1}^a p_i^2 - \sum_{i=1}^{a-1} \sum_{j=i+1}^a 2 (p_i p_j)^2$$

Programa

$$PIC[p_] := 1 - \text{Sum}[p[[i]]^2, \{i, 1, \text{Length}[p]\}] - \text{Sum}[(1/2) * (2 p[[i]] p[[j]])^2, \{i, 1, \text{Length}[p] - 1\}, \{j, i + 1, \text{Length}[p]\}]$$

Ejemplos

```
PIC[ {.1, .9} ]
```

```
0.1638
```

```
PIC[ {.5, .5} ]
```

```
0.375
```

```
PIC[ {.25, .25, .25, .25} ]
```

```
0.703125
```

El origen del PIC reside en el caso de un individuo heterocigoto (X) para una enfermedad dominante. Se busca entonces la probabilidad de que ese locus pueda ser estudiado por medio de un marcador codominante.

El PIC tiene un máximo cuando todos los alelos tienen la misma frecuencia ( $1/a$ )

```
Simplify[pic /. {pi → (1/a), pj → (1/a)}]
```

$$\frac{(-1 + a)^2 (1 + a)}{a^3}$$

El PIC máximo es:

$$M\text{-PIC} = \frac{(a-1)^2 (a+1)}{a^3}$$

---

## Heterocigosidad

La heterocigosidad es una medida más básica de grado de polimorfismo. Es sencillamente la frecuencia esperada de heterocigotos en una población de apareamiento aleatorio.

$$H = 1 - \sum_{i=1}^a p_i^2$$

El estimador de máxima verosimilitud de H es

$$\hat{H} = 1 - \sum_{i=1}^a \hat{p}_i^2$$

Una característica de este MLE es que es sesgado.

Un estimador insesgado de H es:

$$\hat{H}_v = \hat{H}n/(n-1)$$

donde n es el número de alelos muestreados, i.e. 2(#individuos)

La varianza del MLE de H es

$$V(\hat{H}) = \frac{2(n-1) \left( (3-2n) \left( \sum_{i=1}^a p_i^2 \right)^2 + \sum_{i=1}^a p_i^2 + 2(n-2) \sum_{i=1}^a p_i^3 \right)}{n^3}$$

La varianza del estimador insesgado de H es

$$V(\hat{H}_v) = V(\hat{H})n^2/(n-1)^2$$

Programas

```
HeterMLE[p_] := 1 - Sum[p[[i]]^2, {i, 1, Length[p]}
```

### Ejemplo

```
HeterMLE[ {.1, .9} ]
```

```
0.18
```

```

VarHeterMLE[p_, n_] :=
  (2 (n - 1) / n^3) ((3 - 2 n) (Sum[p[[i]]^2, {i, 1, Length[p]}])^2 + 2 (n - 2)
    Sum[p[[i]]^3, {i, 1, Length[p]}] + Sum[p[[i]]^2, {i, 1, Length[p]}])

```

### Ejemplo

```
VarHeterMLE[ {.9, .1}, 1000]
```

```
0.000230004
```

```
HeterUnbiased[p_, n_] := HeterMLE[p] n / (n - 1)
```

### Ejemplo

```
HeterUnbiased[ {.1, .9}, 20]
```

```
0.189474
```

```
VarHeterUnbiased[p_, n_] := VarHeterMLE[p, n] (n^2) / (n - 1)^2
```

### Ejemplo

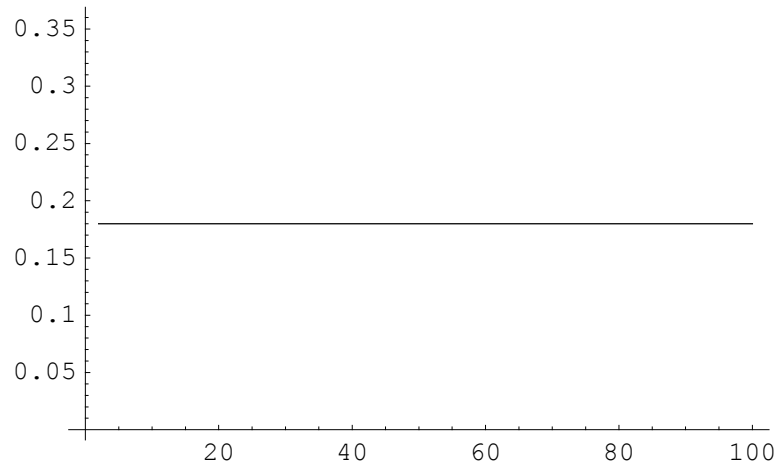
```
VarHeterUnbiased[ {.9, .1}, 20]
```

```
0.0116905
```

## Gráficas

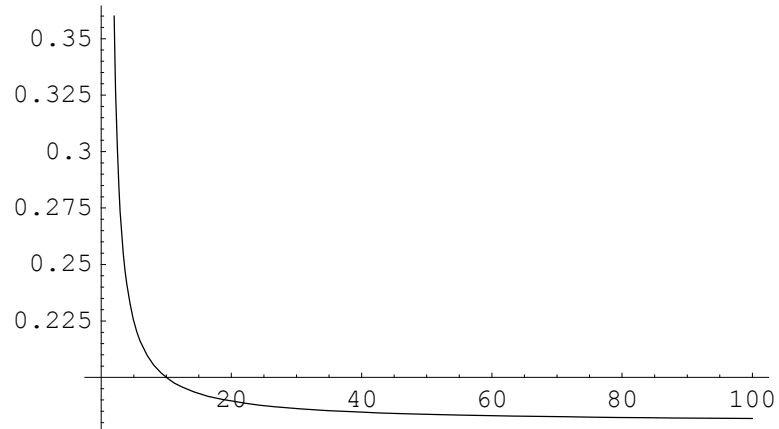
Comparación de los dos estimadores para diferentes tamaños de muestra

```
plot1 = Plot[HeterMLE[ {.1, .9} ], {n, 2, 100}]
```



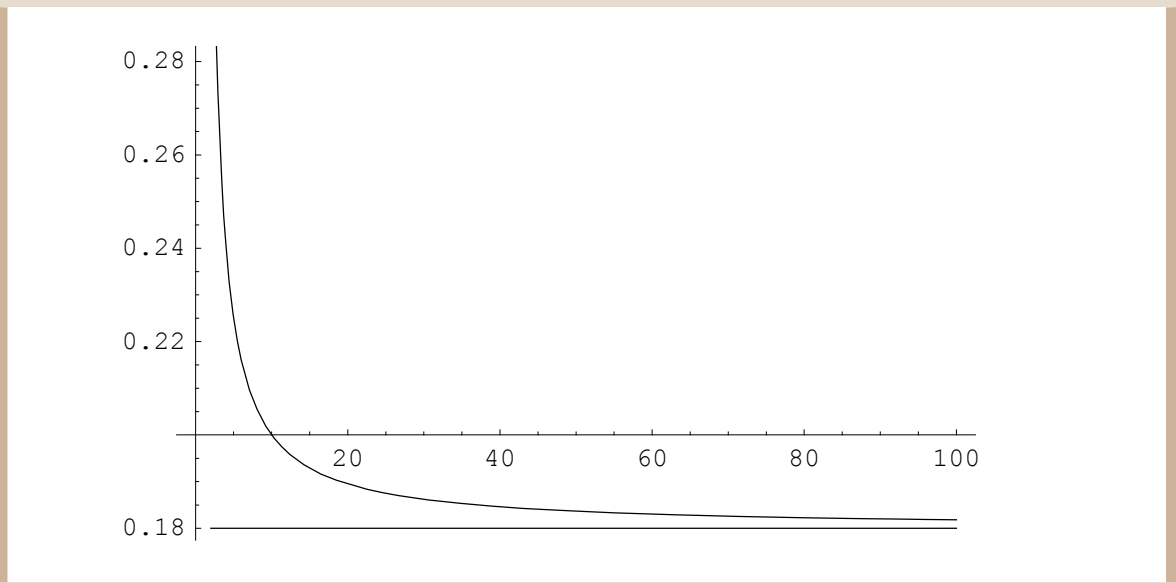
- Graphics -

```
plot2 = Plot[HeterUnbiased[ {.1, .9}, n ], {n, 2, 100}]
```



- Graphics -

```
Show[plot1, plot2]
```



- Graphics -