

# El Modelo Bayesiano

## La Información

- ▶ Los datos se recogen de cada casilla en la muestra y constituyen un vector:  $(X_{ij1}, X_{ij2}, X_{ij3}, X_{ij4}, X_{ij5}, X_{ij6})$
- ▶ Para la casilla  $j$ , del estrato  $i$ , se tiene
  - ▶  $X_{ijk}$  = votos por el  $k$ -ésimo partido (o coalición);  $k = 1, \dots, 4$
  - ▶  $X_{ij5}$  = votos a favor de candidatos no registrados.
  - ▶  $X_{ij6}$  = votos nulos.
  - ▶ La suma es el total de votos emitidos en la casilla.

## Modelo

- ▶ Los datos  $(X_{ij1}, X_{ij2}, X_{ij3}, X_{ij4}, X_{ij5}, X_{ij6})$  se consideran una observación de una distribución Mutinomial con:
  - ▶ Parámetro  $m_{ij}$  (el número de votantes en el listado nominal de la casilla) y
  - ▶ Probabilidades  $(\theta_{ij1}, \theta_{ij2}, \theta_{ij3}, \theta_{ij4}, \theta_{ij5}, \theta_{ij6})$ .

En promedio,  $m_{ij} \approx 550$

Valor máximo  $m_{ij} = 750$

## Modelo

- ▶ Con ese modelo:
  - ▶ La distribución de  $X_{ijk}$  considera Binomial con parámetros  $m_{ij}$  y  $\theta_{ijk}$  (valores de  $m_{ij}$  grandes).
  - ▶ Se aproxima con una distribución Normal.
  - ▶  $\theta_{ijk}$  es homogénea ( $\theta_{ik}$ ) a lo largo de todo un estrato.

## Modelo

- ▶ La esperanza depende de la constante conocida  $m_{ij}$  y del parámetro desconocido  $\theta_{ik}$  ( la proporción de voto a favor del partido  $k$  en el estrato  $i$  ).
- ▶ La varianza depende de las mismas componentes.
- ▶ Es un modelo Normal con un solo parámetro libre (  $\theta_{ik}$  ) para ajustarse a los datos.

## Modelo

- ▶ Con el propósito de aumentar la flexibilidad y la capacidad descriptiva del modelo:
  - ▶ Se adopta para  $X_{ijk}$  una distribución Normal con parámetros:
  - ▶  $\mu_{ijk} = m_{ij} \theta_{ik}$  y  $\sigma_{ijk}^2 = m_{ij} \tau_{ik}^2$
  - ▶  $\theta_{ik} \in (0, 1)$  y  $\tau_{ik} > 0$ , parámetros libres (no relacionados).

## Análisis

- ▶ En este modelo:
  - ▶  $\theta_{ik}$  representa la proporción de personas en el listado nominal (del estrato  $i$ ) que votaron y además lo hicieron a favor del partido  $k$ .
  - ▶ Con una distribución inicial  $P(\theta_{ik}, \tau_{ik}^2)$ , la muestra de casillas del estrato  $i$  ( $D_i$ ), produce la distribución final  $P(\theta_{ik}, \tau_{ik}^2 | D_i)$ .
  - ▶ A partir de  $P(\theta_{ik}, \tau_{ik}^2 | D_i)$ , se obtiene  $P(\theta_{ik} | D_i)$ .

## Análisis

- ▶  $m_{ij}$  es el número de personas en el listado nominal de la casilla  $j$  en el estrato  $i$ . Entonces,  $M_i = \sum_{j=1}^{n_i} m_{ij}$  es número de personas en el listado nominal del estrato  $i$  completo.
- ▶ La distribución posterior de  $V_{ik} = M_i \theta_{ik}$  (el número votos en favor del partido  $k$  en el estrato  $i$ ) es inmediata a partir de  $P(\theta_{ik} | D_i)$ .
- ▶ La distribución de  $V_i = \sum_{k=1}^6 M_i \theta_{ik} = M_i \sum_{k=1}^6 \theta_{ik}$  (el número total de votos emitidos en el estrato  $i$ ) requiere la final conjunta de  $\theta_{ik}$ ;  $i = 1, \dots, 6$ .

## Análisis

- ▶ Los estratos constituyen subpoblaciones que se muestrean en forma independiente y en esas condiciones se obtiene:
  - ▶ La distribución de  $VT_k = \sum_{i=1}^K V_{ik}$  (el número total de votos en favor del candidato k en el país).
  - ▶ La distribución de  $VT = \sum_{i=1}^K V_i$  (el número total de votos emitidos en el país).
  - ▶ La distribución de  $\theta_k = VT_k / VT$  (la proporción de votos emitidos a favor del candidato k en todo el país),  $k = 1, \dots, 4$ .

## Distribución Inicial

- ▶ La distribución inicial para los parámetros del modelo es neutra en tanto que es la misma para todos los candidatos.
- ▶  $X_{ijk} \sim N( X_{ijk} | m_{ij} \theta_{ik}, m_{ij} \tau_{ik}^2 )$  independientes.
- ▶ Iniciales independientes  $P( \theta_{ik}, \tau_{ik}^{-2} )$  Normal-Gamma.
- ▶ Iniciales Truncadas:  $\theta_{ik} \geq 0; ( \theta_{i1} + \theta_{i2} + \theta_{i3} ) \leq 1$

## Procesamiento de la información

- ▶ La información muestral se recibió, en forma de remesas, en las instalaciones del Comité, a intervalos de cinco minutos.
  - ▶ La información muestral se acumuló progresivamente, cubriendo los distintos estratos.
  - ▶ Las distribuciones finales no tienen una expresión analítica y se aproximaron con métodos de simulación.
  - ▶ Cada cinco minutos se llevó a cabo un ciclo completo de cálculo para obtener los resultados de la remesa respectiva.