

Comité Técnico Asesor para el Conteo Rápido

Expresiones matemáticas
para la estimación de la conformación de la Cámara
de Diputados

Estimación de la conformación de la Cámara de Diputados

En los trabajos de determinación de tamaño de muestra, en las simulaciones de evaluación de métodos para estimar intervalos de confianza y en el segundo simulacro el Comité Técnico Asesor para el Conteo Rápido (COTECORA) probó distintas posibilidades a fin de realizar estimaciones de la votación.

Con base en lo anterior, se estableció que, con el enfoque clásico, se emplearan dos tipos de estimadores el común o clásico y el robusto. En tanto que, con el enfoque bayesiano, se trabajará con dos escenarios: el primero considera que el total de votos emitidos en cada casilla tienen una distribución a nivel de distrito, normal multivariada con parámetros desconocidos, el segundo considera una transformación a log-momios de los votos para cada partido, supone que son aproximadamente normales en cada distrito con parámetros desconocidos.

Los detalles de las expresiones matemáticas de los estimadores se describen a continuación:

1. Enfoque Clásico

Propuesta de Patricia Isabel Romero Mares y Adriana Margarita Ducoing Watty

Se selecciona una muestra aleatoria simple sin reemplazo de casillas de cada distrito electoral. Para la obtención de los Diputados por Mayoría, cada distrito electoral se considera como una subpoblación y para la obtención de Diputados por Representación Proporcional se considera a cada distrito electoral como un estrato.

En cada remesa se calcula el estimador clásico del total de votos para cada partido político, coaliciones, candidatos independientes, votos a candidatos no registrados y votos nulos en cada distrito, de la siguiente manera:

$$\hat{Y}_{hi} = \frac{N_h}{n_h} \sum_{j=1}^{n_h} y_{hij},$$

donde:

\hat{Y}_{hi} es el total de votos estimado en el distrito h para la fuerza política i ,

N_h es el total de casillas electorales en el distrito h ,

n_h es el número de casillas electorales en muestra que se observaron en la remesa y

y_{hij} es el número de votos de la casilla electoral j del distrito h para la fuerza política i .

El factor de expansión $\frac{N_h}{n_h}$ se va ajustando conforme se observan en la remesa más casillas del distrito.

Una vez obtenidos estos totales se procede a distribuir los votos de las coaliciones entre sus correspondientes partidos políticos.

La fuerza política o candidato independiente que tenga mayor número de votos es la que gana el distrito

Para estimar el total de votos a nivel nacional, que corresponden a cada contendiente se utilizan los procedimientos clásicos de estimación para un muestreo estratificado aleatorio, es decir, se calcula

$$\hat{Y}_i = \sum_{h=1}^{300} \hat{Y}_{hi}$$

Con estos totales de votos nacionales estimados, se calculan la votación nacional y la válida emitida y aplicando lo establecido en la Ley General de Instituciones y Procedimientos Electorales (LGIPE), se estiman los 200 Diputados de Representación Proporcional.

Para construir intervalos de confianza para el número de Diputados que corresponde a cada partido político se utilizan técnicas de remuestreo,

específicamente el método conocido como Bootstrap, con la variante de Reescalamiento de Rao y Wu (1988). Se obtiene un número grande de muestras con reemplazo de la muestra original (entre 500 y 1000) **respetando el diseño de muestra** utilizado y con tamaño de muestra igual al tamaño de muestra original del estrato menos uno, es decir, $n_h^* = n_h - 1$. Con cada una de estas muestras con reemplazo se estima la configuración de la cámara. A partir de estas repeticiones se obtiene una estimación de la distribución del número total de Diputados (mayoría relativa más representación proporcional) que le corresponde a cada partido político y candidatos independientes, y de ahí, determinando los cuantiles 2.5% y 97.5% se obtienen los intervalos aproximados del 95% de confianza del número total de Diputados para cada uno de los partidos.

Propuesta Carlos Erwin Rodríguez Hernández-Vela y Antonio Soriano Flores

La estrategia para estimar la conformación de la Cámara de Diputados se basa en la técnica de remuestreo bootstrap. La cual consiste en:

- a) Seleccionar una submuestra de la muestra total.
- b) Usar la submuestra para estimar el número total de votos por partido (en cada distrito).
- c) Usar la estimación y seleccionar a los partidos con mayor número de votos en los 300 distritos, de esta manera se obtienen los 300 Diputados por Mayoría Relativa.
- d) Combinar la estimación por distrito y estimar la votación nacional por partido.
- e) Usar las reglas de la LGIPE para obtener los 200 Diputados por Representación Proporcional.

Esto se repite un número grande de veces.

El punto b) requiere de estimar el número total de votos por distrito para cada partido, aquí se han probado los estimadores

Estimador común

Estimador robusto

Estimador de razón

$$\hat{Y}_{jh} = N_h \hat{y}_{jh} \qquad \hat{Y}_{jh} = N_h \hat{y}_{.20jh} \qquad \hat{Y}_{jh} = LN_h \frac{\sum_{i \in S_h} y_{ijh}}{\widehat{LN}_{jh}}$$

En donde:

N_h	Número total de casillas instaladas en el distrito h
n_h	Número de casillas seleccionadas en muestra (recuperadas) en el distrito h
$\hat{y}_{jh} = \frac{\sum_{i \in S_h} y_{ijh}}{n_h}$	Promedio simple de votos para el partido j en el distrito h , tomando como base las casillas recuperadas
y_{ijh}	Número de votos en la i -ésima casilla para el partido j en el distrito h
S_h	Sub-muestra de índices derivado del proceso de remuestreo, i.e. el conjunto S_h se obtiene del conjunto $\{1, 2, 3, \dots, n_h\}$ y como el remuestreo involucra reemplazo pueda haber índices repetidos
$\hat{y}_{.20jh}$	Media truncada o trimmed mean, se calcula de manera similar al promedio simple, sin embargo la muestra se ordena y se descarta el 20% de las observaciones más pequeñas y el 20% de las observaciones más grandes. Es una medida robusta de tendencia central.
LN_h	La Lista Nominal para el distrito h
\widehat{LN}_h	Estimador de la Lista Nominal mediante la muestra recuperada, calculado mediante el estimador común, i.e. promedio de Lista Nominal en las casillas recuperadas en el distrito h , multiplicado por el número de casillas instaladas (N_h)

Mediante ejercicios de simulación se observó que el estimador de razón y el común tienen el mismo desempeño (para nuestro problema particular); sin embargo, el estimador de razón conlleva un problema: las casillas especiales no tienen lista nominal y esto dificulta su aplicación.

En votaciones en donde no hay gran varianza por distrito, el estimador común y el robusto arrojan resultados muy similares, pero cuando la varianza es grande el estimador robusto entrega mejores resultados.

Se decidió utilizar el estimador común y el robusto, de esta manera se tendrá una manera de observar la variabilidad en la votación por distrito.

Propuesta Alberto Castillo Morales, Carlos Hernández Garciadiego y Gabriel Núñez Antonio

Para estimar la conformación de la Cámara de Diputados esta propuesta considera un diseño de muestreo estratificado por distrito, con muestreo aleatorio sin reemplazo de casillas dentro de cada distrito y con tamaño de muestra de 30 casillas por distrito. El tamaño de muestra propuesto es de 9,000 casillas con sobre muestra de 30 casillas en cada distrito de las dos entidades (Baja California y Sonora) que tienen huso horario dos horas diferente al del centro del país.

Para la asignación de escaños de Mayoría Relativa se considera el empleo de métodos de selección del mayor total (de votos para cada partido) basado en procedimientos de comparaciones múltiples (ver, por ejemplo, García-Villalpando, 2000). Este tipo de procedimientos contempla la posibilidad de tener empates en la asignación del mayor total derivado de la información muestral disponible, pero asume ciertos supuestos como normalidad e independencia; los cuales no se cumplen dentro de cada distrito, por lo que se consideró una variante del procedimiento anterior considerando los partidos con posibilidad de empatar. También, se analiza el uso de métodos de remuestreo y/o simulación para modelar este tipo de resultados (empates).

Para el caso de escaños por Representación Proporcional, se emplearán procedimientos clásicos estándar para la estimación del total de votos de cada partido, considerando las ponderaciones dadas por el número de casillas en la muestra con respecto al número de casillas en el estrato. Con los totales de votos estimados para cada partido, se emplearán las reglas de asignación de escaños establecidas en la Ley General de Instituciones y Procedimientos Electorales, tomando en cuenta de manera especial el proceso especificado para la sobre representación.

Derivado de las reglas de asignación tanto para diputados de Mayoría Relativa como para los de Representación Proporcional, el empleo directo de estimadores clásicos tradicionales y su análisis (propiedades) resultan no triviales, especialmente en lo que se refiere a la estimación de la varianza asociada. Dado lo anterior, una vez que se tiene una metodología inicial de la conformación de escaños de la Cámara de Diputados, se emplearán procedimientos de remuestreo dentro de cada distrito electoral como el *Bootstrap no-paramétrico* (ver, por ejemplo, Efron 1979) para obtener una estimación de la variabilidad y estar en condiciones de ofrecer intervalos de confianza que cumplan con los lineamientos correspondientes.

Ecuaciones utilizadas:

Notación:

300 distritos, identificados por el subíndice d : $d=1, \dots, 300$. En el distrito d hay N_d casillas con una Lista Nominal $LN_{d,c}$: $d=1, \dots, 300$, $c=1, \dots, N_d$.

La variable de interés es el número de votos Y . Identificando a la casilla con el subíndice c , al partido con el subíndice p y al distrito con el subíndice d , se tiene que $Y_{p,d,c}$ es el total de votos del partido p en la casilla c del distrito d : $d=1, \dots, 300$, $c=1, \dots, N_d$ y $p=1, \dots, 12$ entendiendo que son 10 partidos y 2 posibles candidatos independientes por distrito; estos últimos casos no tendrán votos si no están definidos para la casilla c en el distrito d .

Se toma una muestra aleatoria con reemplazo de 30 casillas en cada distrito. En la remesa j ($j=1, \dots, k$) llegarán n_d casillas para el distrito d ; n_d es un entero entre 0 y 30 (puede ser mayor que 30 en los distritos con sobre muestra).

La remesa j tiene los datos $y_{p,d,c}$: $d=1, \dots, 300$, $c=1, \dots, n_d$ y $p=1, \dots, 12$, adicionalmente se tienen los datos de votos para candidatos no registrados y votos nulos.

Una vez que se hace el reparto de votos de las coaliciones a los partidos en cada distrito, se determina el conjunto de partidos ganadores g_d , que está formado por los partidos que son estadísticamente indistinguibles del partido con mayor número de votos de acuerdo con

$$g_d = \{p \mid y_{\max,d} - y_{p,d} < C_{0.95} V^{1/2} (y_{\max,d} - y_{p,d})\}$$

Donde:

$$y_{\max,d} = \max_{p=1,\dots,12} \{y_{p,d}\}$$

$C_{0.95}$ = percentil correspondiente del 95%,

$V^{1/2} (y_{\max,d,c} - y_{p,d,c})$ = desviación estándar de la diferencia de votos por distrito del partido que obtuvo el máximo y otro de los 11 partidos o candidatos independientes si los hay.

Para las estimaciones del total de votos de cada partido $y_{p,total}$; $p=1,\dots,10$ se utiliza la ponderación N_d/n_d , esto es

$$y_{p,total} = \sum_{d=1}^{300} \sum_{c=1}^{n_d} \frac{N_d}{n_d} y_{p,d,c}; \text{ para } p=1,\dots,12,$$

La misma ecuación se emplea para estimar el total de votos.

Con las estimaciones se lleva a cabo la repartición de diputados siguiendo las reglas de la Ley General de Instituciones y Procedimientos Electorales, y utilizando remuestreo en cada uno de los distritos se estima la variabilidad, para poder construir los correspondientes intervalos de confianza tanto para el número de Diputados de Representación como para la proporción de votos que corresponde a cada partido y el porcentaje de participación.

2. Enfoque Bayesiano

Propuesta de Raúl Rueda Díaz del Campo e Ilan Morgenstern Kaplan

El objetivo del Conteo Rápido es obtener el día 7 de junio, a pocas horas del cierre de las casillas electorales, una estimación confiable de la distribución de la Cámara de Diputados. Como es sabido, la Cámara está conformada por 500 Diputados, 300 elegidos por el Principio de Mayoría Relativa en cada uno de los distritos electorales federales que conforman el país y 200 por el Principio de Representación Proporcional, de acuerdo a la fracción II del artículo 54 de la

Constitución y los artículos 15.1, 15.2, 16 y 17 de la Ley General de Instituciones y Procedimientos Electorales.

Para alcanzar dicho objetivo es relevante estimar, en cada distrito, la votación válida emitida. El proceso que proponemos, se describe a continuación.

Supongamos que el *i*-ésimo distrito tiene M_i casillas electorales y que una muestra de m_i de ellas, es obtenida. En cada casilla observamos un vector de dimensión p , que incluye los votos emitidos a favor de todos los partidos, las coaliciones, los candidatos independientes, los votos nulos y los candidatos no registrados. Es decir, en cada casilla de la muestra, tenemos a $\mathbf{x}_{ij} = (x_{ij1}, \dots, x_{ijp})$, en donde x_{ijk} son los votos emitidos por el partido k en la casilla j del distrito i . De esta manera, el vector con el total de votos emitidos a cada partido en el distrito i , digamos \mathbf{X}_i , puede ser escrito como

$$(1) \quad \mathbf{X}_i = m_i \bar{\mathbf{x}}_i + (M_i - m_i) \bar{\mathbf{X}}_i^*$$

en donde $\bar{\mathbf{x}}_i$ y $\bar{\mathbf{X}}_i^*$ son la media aritmética de la muestra observada y la media aritmética de las casillas no observadas. Así, estimar \mathbf{X}_i es equivalente a estimar $\bar{\mathbf{X}}_i^*$.

Desde la perspectiva bayesiana, toda incertidumbre debe ser cuantificada por una medida de probabilidad. Esto lleva a considerar un modelo que describa el comportamiento de la votación a nivel distrital y actualizarlo con la información que proporciona una muestra aleatoria de casillas. La idea básica es modelar el total de votos emitidos para cada partido a nivel distrital o alguna transformación adecuada. En lugar de hacerlo directamente, se hace en dos pasos: un modelo paramétrico y una distribución inicial sobre los parámetros desconocidos.

Dadas las características del ejercicio, las conclusiones deben estar basadas fundamentalmente en la información muestral, por lo que se usarán distribuciones iniciales poco informativas. El objetivo final es encontrar la distribución predictiva final del total de votos emitidos. Con ella, se simula un número razonablemente grande de posibles configuraciones de la Cámara y con ella, se obtienen intervalos de confianza (en realidad, de probabilidad) para cada uno de los partidos contendientes (incluyendo, candidatos independientes). Además, pueden encontrarse intervalos de probabilidad para la proporción de votación esperada para cada partido y para la participación.

El proceso que proponemos consiste en suponer que X_i o alguna transformación de ella, tiene una distribución normal p -variada con parámetros (\mathbf{m}, H) desconocidos, a los que les asignamos una distribución no informativa de la forma $\pi(\mathbf{m}, H) \propto |H|^{-v/2}$ y que nos lleva a una distribución final normal-Wishart. Con esta distribución es relativamente sencillo demostrar que la distribución predictiva de \bar{X}_i^* es una distribución Student, de la forma

$$Stu_p(\bar{X}_i^* | m_i - v + 1, \bar{x}_i, (m_i - v + 1)k_i S_i^{-1})$$

en donde k_i es una función de M_i y m_i y que depende de la transformación usada en la suposición de normalidad. S_i es la matriz de varianzas y covarianzas muestral en el distrito i .

Para cada distrito se genera una muestra de tamaño n y se inserta en la ecuación (1) para estimar los votos emitidos para cada partido y elegir al ganador en cada distrito. Las votaciones distritales sirven para estimar la votación nacional y determinar cuáles partidos alcanzan el registro. Con los partidos con registro, se procede a repartir los 200 escaños restantes, teniendo así una muestra de tamaño n de la distribución final predictiva de la configuración de la Cámara. Con ella, se obtienen los intervalos de probabilidad 0.95, ya sea para todos los partidos o los intervalos marginales para cada partido.