

Estimación del sub-reporte de casos activos debido al rezago en el reporte de casos: Metodología de Donker

Este algoritmo fue utilizado por Donker et al. (2011), para estimar la cantidad de muertes por Influenza en tiempo real en los Países Bajos en el año 2009. Esta es una adaptación para los infectados por covid-19 en Chile.

1. Metodología

Se necesita contar con el número de pacientes contagiados en cada semana del estudio. El retraso en el aviso de Influenza se midió como la diferencia entre la semana de admisión de un paciente y la semana del informe del virus. Se deben observar todos los retrasos en los informes para todos los casos desde la fecha de interés con el fin de medir la Distribución de estos retrasos. La función de Distribución de frecuencia acumulada entrega la probabilidad de que un caso de covid haya sido i semanas luego del día en que el paciente empezó a mostrar síntomas. Esta probabilidad se definirá como ρ_i .

Luego, se estima el real número de casos. La semana en marcha se ajusta como 0. La meta es estimar el número de contagios i semanas atrás, N_i . Se denota como el número de casos que hayan sido informados con COVID-19 positivo desde aquella semana hasta la semana en marcha como C_i . Notamos entonces, que el número de casos observados es el resultado del producto entre la probabilidad de que el caso haya sido informado y el real número de admisiones N_i . Esto es:

$$C_i = \rho_i N_i$$

Reordenando:

$$\hat{N}_i = \frac{C_i}{\rho_i}. \quad (1)$$

2. Estimador máximo verosímil

Para construir una función de probabilidad para el real número de casos suponemos que el número de casos observados sigue una Distribución binomial que se define por un número de ensayos independientes de N_i , donde la probabilidad de el éxito es ρ_i . La probabilidad de observar C_i casos es,

$$P(C_i|N_i, \rho_i) = \binom{N_i}{C_i} \rho_i^{C_i} (1 - \rho_i)^{N_i - C_i}. \quad (2)$$

La función de verosimilitud correspondiente para N_i dado el número de casos observados C_i y probabilidad de informar ρ_i está dada por,

$$L(N_i; C_i, \rho_i) = \binom{N_i}{C_i} (1 - \rho)^{N_i - C_i}. \quad (3)$$

De esta forma, el valor del real número de casos que maximiza el valor de la función de verosimilitud es

$$\hat{N}_i = \frac{C_i}{\rho_i}. \quad (4)$$

Esto confirma que el estimador directo calculado anteriormente es un estimador de máxima verosimilitud.

Referencias

Donker, T., van boven, M., van Ballegooijen, M., Schurink-van 't Klooster, T., Wielders, C., and Wallinga, J. (2011). Nowcasting pandemic influenza a/h1n1 2009 hospitalizations in the netherlands. *European journal of epidemiology*, 26:195–201.