ESTIMACIÓN DE MÁXIMA VEROSIMILITUD

La clase anterior vimos que la estimación Bayesiana une creencias con datos. Su ventaja es que nos permite ir actualizando lo que pensamos sobre un estimador. Este proceso Bayesiano es cercano al pensamiento humano ya que muchas veces cuando tenemos una idea sobre algo y los datos muestran otra cosa, no cambiaremos totalmente nuestra idea, sino que solo la ajustamos un poco. Sin embargo, matemáticamente Bayes tiene algunas dificultades, siendo una de las más importantes la elección de la distribución a priori. ¿Cómo saber qué forma matemática tiene mi creencia? ¿Por qué elegir determinada f.d.p y no otra? Este tipo de desafíos son objeto de grandes discusiones entre estadísticos. Hasta hace no mucho tiempo, también el cálculo computacional de los modelos Bayesianos era un desafío complejo. Por estas y otras razones, el método más frecuentemente utilizado para obtener estimadores es máxima verosimilitud.

Los estimadores de máxima verosimilitud se basan solamente en la muestra, sin tomar en cuenta nuestras creencias. Por lo tanto, no hacen uso de distribuciones a priori o a posteriori, sino que solamente usan la función de verosimilitud (de ahí su nombre). Como vimos anteriormente, la función de verosimilitud no es más que la función que engloba la probabilidad de toda la muestra.

Para lograr encontrar el estimador utilizando este método, se asume que la muestra es representativa de toda una población. Por lo tanto, lo que se quiere es que la probabilidad de la muestra sea lo más alta posible. Por lo tanto, se maximiza la función de verosimilitud según el parámetro que deseamos estimar. Matemáticamente, se deriva la función de verosimilitud con respecto al parámetro y se iguala a cero. A partir de esto se puede despejar el parámetro estimado.

Ejemplo Máxima Verosimilitud

Tenemos una muestra de la cantidad de personas que entran a un museo el día sábado entre las 12 y 1pm. Nos interesa estimar cuánto es la probabilidad de que a esa hora entren más de 1000 personas. El director del museo quiere conseguir esta estimación para decidir a cuántos trabajadores necesita contratar como guías en ese horario.

Este requerimiento puede ser formulado en lenguaje estadístico. Diríamos que tenemos una variable X que representa la cantidad de personas que entran al museo entre las 12 y 13pm del día sábado. Al ser una variable del tipo “número de eventos en un ciclo” sabemos tiene una distribución Poisson:

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Es decir, la **función de probabilidad** de x, dado un parámetro lambda, es igual a la división entre exponencial elevado a menos lambda por lambda elevado a x, todo esto partido por x factorial.

Para poder responder al requerimiento del director del museo necesitamos responder a la siguiente pregunta: ¿cuánto vale λ? Como no conocemos el valor de lambda, tendremos que estimarlo. En este ejemplo veremos cómo realizar la estimación de lambda utilizando el método de máxima verosimilitud.

Lo primero es asumir que en nuestra muestra todas las observaciones siguen la misma distribución. Luego, para crear nuestra función de verosimilitud asumimos que todas las observaciones son independientes unas de otras. Por lo tanto, si queremos la probabilidad conjunta de toda la muestra, tendremos que **multiplicar la función de distribución de cada probabilidad de la siguiente forma**:

$$f(\boldsymbol{x}|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Es decir, la **función de verosimilitud** es igual a la pitatoria de la función de probabilidad individual. Si desarrollamos lo anterior usando las **propiedades de los productos y potencias**, tendremos lo siguiente:

$$L(\lambda) = f(\boldsymbol{x}|\lambda) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Esta función, por lo general la denotamos como L, ya que viene de Likelihood, que es la traducción de verosimilitud al inglés. Ahora que tenemos la función, tendríamos que maximizarla, ya que queremos que la probabilidad sea lo más alta posible. Por lo tanto, tendríamos que derivar la función con respecto a λ. Sin embargo, la función anterior es muy compleja de derivar. Por lo tanto, lo que hacemos primero es usar la **función logaritmo**:

$$l(\lambda) = \ln(L(\lambda)) = \ln\left(\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x\_i!}\right)$$

En esta fórmula denotamos con una L minúscula (l) el logaritmo de la verosimilitud, que en inglés es conocido como el log-likelihood. Si usamos las **propiedades del logaritmo**, llegaremos a:

$$l(\lambda) = -n\lambda + \ln\lambda \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

Esta expresión es mucho más fácil de derivar, por lo tanto, tomamos derivadas con respecto al parámetro:

$$\frac{\partial l(\lambda)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

Para encontrar un máximo tenemos que igualar la derivada anterior al promedio y luego despejar nuestro estimador:

$$\begin{aligned} -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i &= 0 \\ \frac{1}{\lambda}\sum_{i=1}^{n} x_i &= n \\ \frac{1}{n}\sum_{i=1}^{n} x_i &= \hat{\lambda} \\ \hat{\lambda} &= \bar{x} \end{aligned}$$

Por lo tanto, **el estimador de máxima verosimilitud para λ (lambda) es el promedio**. Eso significa que, por ejemplo, si en nuestra muestra en promedio entran 800 personas al museo entre las 12 y las 1pm del sábado, entonces un buen estimador de máxima verosimilitud para λ será 800.

Ejercicio 1

El entrenador de basquetbol de un equipo amateur quiere saber quién debe ingresar para lanzar el último tiro en los últimos 5 segundos del encuentro. Entonces quiere saber la probabilidad de éxito de su jugador estrella.
¿Cuál es el estimador de máxima verosimilitud para el parámetro p de una distribución Bernoulli del jugador?

Invarianza del Estimador MV

Una de las propiedades más útiles de un estimador de máxima verosimilitud es la invarianza. Esta propiedad nos dice que, **si quiero encontrar un estimador para una función de un parámetro, el estimador será la misma función evaluada en el estimador**.

Como estudiaron en estadística descriptiva, para cada distribución tenemos una fórmula que nos permite estimar la mediana, media, varianza y otros estadísticos. Por ejemplo, supongamos que tengo una muestra que sigue una distribución Exponencial y que estoy interesado en conocer la mediana de esta muestra para así poder generar dos grupos. ¿Cómo puedo estimarla?

Lo primero, es recordar cómo es la distribución Exponencial:

$$f(x|\lambda) = \lambda \exp(-\lambda \cdot x)$$

Es decir, un parámetro lambda, por la exponencial de menos lambda por x. La mediana de esta función, que es exponencial, la representaremos por m y tiene la siguiente forma:

$$m = \frac{\ln(2)}{\lambda}$$

Es decir, la mediana es igual al logaritmo natural de dos dividido por lambda. ¿Cómo estimamos m? Según lo que hemos aprendido hasta ahora, tendríamos que encontrar una distribución donde estuviese m para así poder maximizarla. Sin embargo, gracias a la propiedad de invarianza podemos decir que si encontramos $\hat{\lambda}$, entonces podemos definir el estimador de m con la siguiente fórmula:

$$\hat{m} = \frac{\ln(2)}{\hat{\lambda}}$$

Por lo tanto, tenemos que preocuparnos solo de obtener el estimador de lambda, para poder encontrar un estimador de m.

Ejercicio 2

Encuentre el estimador de lambda para esta distribución exponencial.

CALCULO EN STATA

Realizar el cálculo en Stata de un estimador por máxima verosimilitud no es trivial. Por lo general, es el investigador quien tiene que escribir la función de verosimilitud que desea maximizar y luego indicarle a Stata que lo haga.

Cambiemos un poco el primer ejemplo de esta clase, sobre la muestra de la cantidad de personas que entran a un mueso entre las 12:00 y las 13:00. Antes estábamos interesados solamente en el día sábado, pero supongamos ahora que la muestra incluye datos para todos los días por seis meses. Es decir, que tenemos aproximadamente 180 datos.

Lo primero que haremos será generar una simulación de datos y fijar la cantidad de observaciones en 180. Luego generamos la variable con distribución Poisson y media poblacional 800. Y luego estimamos el promedio de personas. El código será el siguiente:

1. Fijamos la 'semilla' para que la simulacion siempre de los mismos valores:

   ```
   set seed 1234
   ```

2. Fijamos la cantidad de observaciones que queremos:

   ```
   set obs 180
   ```

3. Generamos la variable, que es Poisson con media poblacional 800:

   ```
   gen personas = rpoisson(800)
   ```

4. El estimador es el promedio:

   ```
   mean personas
   ```

Si realizamos el procedimiento anterior de forma correcta, el resultado será:

```
Mean estimation                   Number of obs    =     180

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
    personas |   801.4833   2.041309      797.4552    805.5115
--------------------------------------------------------------
```

El promedio de la variable, lo que es lo mismo que el estimador lambda para esta distribución Poisson, tiene un valor de aproximadamente 801. Este ejercicio nos muestra que cuando la muestra original es una Poisson con parámetro poblacional 800, el método de máxima verosimilitud logra encontrar un valor muy cercano a este.