Un cambio de enfoque: De frecuentista (pre-experimental) a bayesiano (pos-experimental)

Alan Ledesma Arista.[1]

[1] alan.ledesma@bcrp.gob.pe

"*From where we stand, the rain seems random.*
If we could stand somewhere else, we would see the order in it."

Desde donde estamos, la lluvia parece aleatoria.
Si pudiéramos pararnos en otro lugar, veríamos el orden en ella.

— T. Hillerman (1990) Coyote Waits.

Tomado de Robert 2007

Material elaborado en base a Herbst y Schorfheide 2016 (cap. 3) y Robert 2007 (cap. 1)

Ejemplo motivación

Suponga que en una de las pruebas de un examen de salud ocupacional, usted resulta positivo para una enfermedad rara degenerativa que afecta al 0,1 % de la población. Debido a las importantes consecuencias en la salud de dicha enfermedad, usted consulta con un médico sobre la posibilidad de un falso positivo.
El médico le informa que la prueba con resultado positivo identifica correctamente al 99 % de personas que tienen la enfermedad. Por lo tanto, sólo hay un 1 % de falsos positivos.

- ¿Cuál es la probabilidad de realmente tener la enfermedad?

Ejemplo motivación

Necesitamos en realidad el teorema de Bayes

- Evento E: tener la enfermedad
- Evento P: contar con una prueba positiva

Teorema de Bayes

$$\mathrm{p}(E|P) = \frac{\mathrm{p}(P|E)\mathrm{p}(E)}{\mathrm{p}(P)} = \frac{\mathrm{p}(P|E)\mathrm{p}(E)}{\mathrm{p}(E)\mathrm{p}(P|E) + \mathrm{p}(\neg E)\mathrm{p}(P|\neg E)}$$

Donde:

- Probabilidad *a posteriori* de tener la enfermedad es $\mathrm{p}(E|P)$

  Esto es lo que queremos calcular
- Probabilidad *a priori* de tener la enfermedad es $\mathrm{p}(E)$

  Usualmente es un "supuesto razonable".
  Por ejemplo, la frecuencia de la enfermedad en la población (i.e., $\mathrm{p}(E) = 0{,}001$)
- La probabilidad de tener una prueba positiva dado que tenemos la enfermedad: $\mathrm{p}(P|E) = 0{,}99$
- La probabilidad de falso positivo: $\mathrm{p}(\neg P|E) = 0{,}01$

Ejemplo motivación

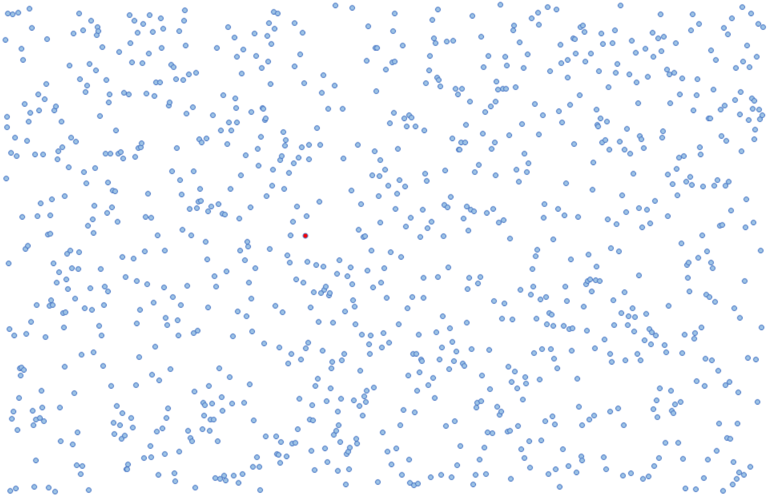
Así, el teorema de Bayes nos lleva a que

$$\mathrm{p}(E|P) = \frac{0{,}99 \times 0{,}001}{0{,}001 \times 0{,}99 + (1 - 0{,}001) \times 0{,}01} = 9\,\%$$

- La probabilidad de tener la enfermedad luego de dar positivo es $\mathrm{p}(E|P) = 9\,\%$ lo que resulta notablemente inferior a lo que anunció el doctor ($\mathrm{p}(P|E) = 99\,\%$)
- ¿Tiene sentido que luego de dar positivo, la probabilidad de tener la enfermedad sea tan baja?
- Esta es una excelente oportunidad para comparar la inferencia pre-experimental contra la pos-experimental!

Ejemplo motivación

- En una muestra de 1000 personas, ●, se espera que una persona tenga la enfermedad, ●



Ejemplo motivación

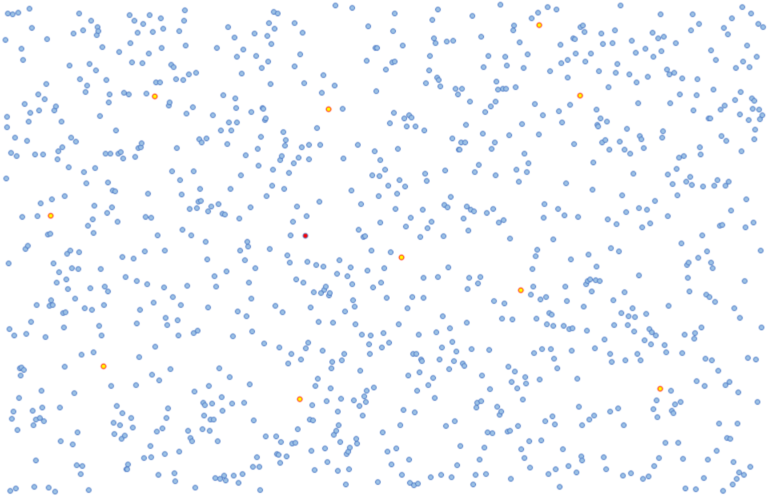
- Con un **enfoque pre-experimental** se tendría que repetir muchas veces el ejercicio con muchas muestras y con ello calcular la probabilidad de dar positivo si se tiene la enfermedad (99 %)
- Con ello la probabilidad de falso positivo es de 1 %

Ejemplo motivación

- Con un **enfoque pos-experimental** se utiliza la información del experimento (la muestra) para mejorar la inferencia sobre la probabilidad de tener la enfermedad

Ejemplo motivación

- Luego de hacer la prueba, se tendrán $1000 \times 1\,\% = 10$ falsos positivos,



Ejemplo motivación

- La '**muestra relevante**' se reduce solo a aquellos que dan positivo

- En esta muestra relevante, la probabilidad de tener la enfermedad es $1/11 \approx 9\,\%$
- I.e., $9\,\%$ es una probabilidad más apropiada por que el evento '**dar positivo a la prueba**' reduce el espacio muestral
- Note que la inferencia pos-experimental no descarta la pre-inferencial ...
- ... por el contrario, la inferencia pos-experimental utiliza los resultados de la pre-inferencial

$\rightarrow$ Si se hace una nueva prueba y esta resulta positiva ¿cual es la probabilidad de tener la enfermedad?

Modos de inferencia

- Inferencia frecuentista: procedimientos de muestreo repetido
- Las medidas de precisión y desempeño de evaluación estadística son "pre-experimentales"
- Sin embargo, muchos estadísticos consideran que se debe utilizar el información "post-experimental" para evaluar la inferencia
- Así, sólo la observación efectiva de los datos es relevante mas no las otras observaciones posibles en el espacio muestral

Modos de inferencia

Suponga que X_1 y X_2 son dos variables aleatorias idénticas e independientemente distribuidas de la siguiente manera

$$p(\{X_i = \theta - 1\}) = p(\{X_i = \theta + 1\}) = \frac{1}{2} \text{ para } i \in \{1, 2\}$$

considere el siguiente intervalo de confianza

$$C(X_1, X_2) = \begin{cases} \frac{1}{2}(X_1 + X_2) & \text{Si } X_1 \neq X_2 \\ X_1 - 1 & \text{Si } X_1 = X_2 \end{cases}$$

- Con una perspectiva pre-experimental $C(\cdot)$ es un intervalo de confianza de $\boxed{}$
- Con una perspectiva post-experimental $C(\cdot)$ es un intervalo de confianza de $\boxed{}$ si $\boxed{}$ y de $\boxed{}$ si $\boxed{}$
-

Modos de inferencia

Suponga que X_1 y X_2 son dos variables aleatorias idénticas e independientemente distribuidas de la siguiente manera

$$p(\{X_i = \theta - 1\}) = p(\{X_i = \theta + 1\}) = \frac{1}{2} \text{ para } i \in \{1, 2\}$$

considere el siguiente intervalo de confianza

$$C(X_1, X_2) = \begin{cases} \frac{1}{2}(X_1 + X_2) & \text{Si } X_1 \neq X_2 \\ X_1 - 1 & \text{Si } X_1 = X_2 \end{cases}$$

- Con una perspectiva pre-experimental $C(\cdot)$ es un intervalo de confianza de $\boxed{75\,\%}$
- Con una perspectiva post-experimental $C(\cdot)$ es un intervalo de confianza de $\boxed{100\,\%}$ si $\boxed{X_1 \neq X_2}$ y de $\boxed{50\,\%}$ si $\boxed{X_1 = X_2}$
- ¿Tiene sentido concentrarse en medidas de precisión pre-experimental, cuando se sabe que puede ser falible después de ver los datos?

Principios de estimación

- **Principio de condicionalidad**: Si un experimento es seleccionado por algún mecanismo aleatorio independiente del parámetro desconocido θ, entonces sólo el experimento realmente realizado es relevante

- **Principio de suficiencia**: Considere un experimento para determinar el valor de un parámetro desconocido θ y suponga que $S(\cdot)$ es un estadístico suficiente. Si $S(X_1) = S(X_2)$, entonces X_1 y X_2 contienen la misma evidencia sobre θ

Ambos principios implican:

- **Principio de verosimilitud**: Toda la información sobre un parámetro desconocido θ que se puede obtener de un experimento está contenida en la probabilidad de θ condicional a los datos.
  Así, dos funciones de verosimilitud para θ (del mismo o de diferentes experimentos) contienen la misma información sobre θ si son proporcionales entre sí.

- En general el método frecuentista no satisface el principio de verosimilitud, en contraste el enfoque bayesiano sí.

Enfoque Bayesiano

- El enfoque parte de creencias de probabilidad (prior distribution), éstas se actualizan con la información en la muestra (verosimilitud) y de este refinamiento se obtiene la estimación de las distribuciones de los parámetros (posterior distribution).
- Sea el PGD caracterizado por parámetros θ con *pdf* $p(\theta)$.

1ro *Prior distribution* sobre $p(\theta)$: $p(\theta)$ (conocido: creencias)

2do *Verosimilitud de la data*: $p(Y|\theta)$ (conocido: Econ. frecuentista)

3ro *Posterior distribution*: $p(\theta|y)$ (desconocido).
De la regla de Bayes:

$$p(Y \cap \theta) = p(Y|\theta)p(\theta) = p(\theta|Y)p(Y) \Rightarrow p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}$$

$$p(\theta|Y) \propto p(Y|\theta)p(\theta)$$

$\Rightarrow \widehat{p}(\theta) = p(\theta|Y)$

- A diferencia del marco frecuentista, θ se considera variable aleatoria
- Sin embargo, esto no implica que θ se determine en un experimento aleatorio
- El cálculo de su probabilidad se utiliza para caracterizar el estado del conocimiento

Enfoque Bayesiano: dificultad técnica

- Note que con el enfoque bayesiano se obtiene $\widehat{p}(\theta) = p(\theta|Y)$ y no $\widehat{\theta}$.
  ¿Cuál es la diferencia?

- Para conocer las cualidades de la estimación de θ hay que estudiar las propiedades de la distribución *posterior*

- Por ejemplo, se pueden conocer los momentos de θ de manera analítica si $p(\theta|Y)$ es alguna PDF conocida o de manera numérica (integración de montecarlo) de caso contrario

- La dificultad es que $p(\theta|Y)$ no suele tener una forma analítica conocida

- Tener la habilidad de simular $p(\theta|Y)$ es central:
  gibbs sampling o *Metropolis-Hasting*

Enfoque Bayesiano: Crítica

- Toda inferencia bayesiana es sensible a la elección de la distribución *prior* $p(\theta)$
- La *prior* es producto de la creencia inicial del investigador y es, por lo tanto, "subjetiva"
- Muchos econometristas creen que los resultados de una investigación no deben depender de creencias subjetivas; y por lo tanto, son muy escépticos del enfoque bayesiano
- Sin embargo, todo análisis implica elecciones subjetivas

Enfoque Bayesiano - Ejemplo: Aplicación simplificada

Sea el espacio paramétrico $\Theta = \{0, 1\}$ y el espacio muestral $\mathcal{Y} = \{0, 1, 2, 3, 4\}$.
La distribución de probabilidad P se define sobre el espacio de producto $\Theta \times \mathcal{Y}$

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$P_{\theta=0}(Y)$	0,750	0,140	0,040	0,037	0,033
$P_{\theta=1}(Y)$	0,700	0,251	0,040	0,005	0,004

Suponga que el prior es $p(\theta = 0) = p(\theta = 1) = 0{,}5$. Entonces, si el experimento se materializa en una observación $Y = 1$, ¿que dice eso sobre θ?

- La densidad marginal de $Y = 1$ es

$$\begin{aligned} P(Y = 1) =& P(Y = 1|\theta = 0)P(\theta = 0) + P(Y = 1|\theta = 1)P(\theta = 1) \\ =& \boxed{} \cdot \boxed{} + \boxed{} \cdot \boxed{} = \boxed{} \end{aligned}$$

- La distribución posterior para $\theta = 0$ o $\theta = 1$ es

$$P(\theta = 0|Y = 1) = \frac{P(Y = 1|\theta = 0)P(\theta = 0)}{P(Y = 1)} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

$$P(\theta = 1|Y = 1) = \frac{P(Y = 1|\theta = 1)P(\theta = 1)}{Y = 1} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

- La observación $Y = 1$ provee evidencia en favor de $\boxed{}$

Enfoque Bayesiano - Ejemplo: Aplicación simplificada

Sea el espacio paramétrico $\Theta = \{0, 1\}$ y el espacio muestral $\mathcal{Y} = \{0, 1, 2, 3, 4\}$.
La distribución de probabilidad P se define sobre el espacio de producto $\Theta \times \mathcal{Y}$

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	$Y = 4$
$P_{\theta=0}(Y)$	0,750	0,140	0,040	0,037	0,033
$P_{\theta=1}(Y)$	0,700	0,251	0,040	0,005	0,004

Suponga que el prior es $p(\theta = 0) = p(\theta = 1) = 0{,}5$. Entonces, si el experimento se materializa en una observación $Y = 1$, ¿que dice eso sobre θ?

- La densidad marginal de $Y = 1$ es

$$\begin{aligned} P(Y = 1) =& P(Y = 1|\theta = 0)P(\theta = 0) + P(Y = 1|\theta = 1)P(\theta = 1) \\ =& \boxed{0{,}14} \cdot \boxed{0{,}5} + \boxed{0{,}251} \cdot \boxed{0{,}5} = \boxed{0{,}1955} \end{aligned}$$

- La distribución posterior para $\theta = 0$ o $\theta = 1$ es

$$P(\theta = 0|Y = 1) = \frac{P(Y = 1|\theta = 0)P(\theta = 0)}{P(Y = 1)} = \frac{\boxed{0{,}07}}{\boxed{0{,}1955}} = \boxed{0{,}358}$$

$$P(\theta = 1|Y = 1) = \frac{P(Y = 1|\theta = 1)P(\theta = 1)}{Y = 1} = \frac{\boxed{0{,}1255}}{\boxed{0{,}1955}} = \boxed{0{,}642}$$

- La observación $Y = 1$ provee evidencia en favor de $\boxed{\theta = 1}$

References I

Robert, Christian P (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Segunda. Springer.

Herbst, Edward P y Frank Schorfheide (2016). *Bayesian estimation of DSGE models*. Princeton University Press.