EST-46115: Modelación Bayesiana

Profesor: Alfredo Garbuno Iñigo — Primavera, 2021.
Objetivo. Repasar notación que utilizaremos a lo largo del curso. Y a la vez, establecer la motivación de los temas que trataremos en la materia.
Lecturas recomendas: Notas del curso de fundamentos (2022) y sección 1 de [2].

1. NOTACIÓN

Usaremos la convención usual en probabilidad.

1.1. Variables aleatorias

Una variable aleatoria X está definida a través de un `espacio de probabilidad` $(\mathcal{X}, \mathcal{F}, \pi)$. La función $\pi : \mathcal{F} \to [0, 1]$ se llama la `función de distribución` de la variable aleatoria X. Escribimos $X \sim \pi$.

1.2. Distribución paramétrica

Decimos que una función de distribución es `paramétrica` si se puede identificar completamente la distribución con respecto a un `vector de parámetros` $\theta \in \mathbb{R}^p$. Esto lo denotamos de la siguiente manera

$$\pi_\theta(x) \qquad \text{ó} \qquad \pi(x; \theta)\,, \tag{1}$$

y si $\theta \neq \theta'$ entonces $\pi_\theta(x) \neq \pi_{\theta'}(x)$ para cualquier x en el `soporte`.

1.3. Valores esperados

El `valor esperado` de una variable aleatoria $X \sim \pi$ se define como

$$\mathbb{E}[X] = \int_{\mathcal{X}} x\, \pi(x)\, \mathrm{d}x\,. \tag{2}$$

1.3.1. Abuso de notación La `función de densidad` está definida como $\mathrm{d}\pi/\mathrm{d}x$. En la definición de valor esperado deberíamos de haber escrito $\mathrm{d}\pi(x)$ o bien $\pi(\mathrm{d}x)$ (integrales de Lebesgue). Pero para no ofuscar notación, y dado que no veremos cosas exóticas en el curso, lo obviamos...

1.4. Estadísticas de interés

La definición se puede extender con $f : \mathcal{X} \to \mathbb{R}$ y se calcula como

$$\mathbb{E}[f(X)] = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x\,. \tag{3}$$

Denotaremos de la siguiente manera

$$\pi(f) := \mathbb{E}[f(X)]\,. \tag{4}$$

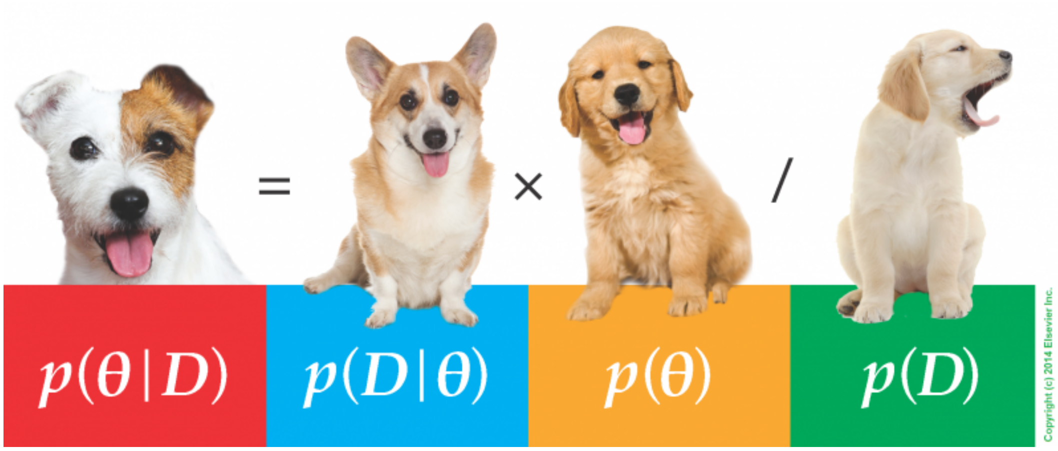


FIGURA 1. *Tomado de [4]* .

1.5. Probabilidad condicional

La `probabilidad condicional` de A dado el evento B se denota $\pi(A|B)$ y está definida como

$$\pi(A|B) = \frac{\pi(A \cap B)}{\pi(B)} \tag{5}$$

2. REPASO INFERENCIA

Repaso de inferencia bajo un enfoque frecuentista.

2.1. Regla de Bayes

La `regla de Bayes` utiliza la definición de probabilidad condicional para hacer inferencia a través de

$$\pi(A|B) = \frac{\pi(B|A)\pi(A)}{\pi(B)} . \tag{6}$$

2.2. Ejemplos

- Verosimilitud: $x|\theta \sim \mathsf{Binomial}(n, \theta)$ + Previa: $\theta \sim \mathsf{Beta}(\alpha, \beta)$ = Posterior: ?
- Verosimilitud: $x|\theta \sim \mathsf{Uniforme}(0, \theta)$ + Previa: $\theta \sim \mathsf{Pareto}(\theta_0)$ = Posterior: ?

3. REPASO INFERENCIA

Repaso de inferencia bajo un enfoque bayesiano.

3.1. Ejemplo

Este ejemplo fue tomado de [3].

3.2. Diferentes previas, diferentes posteriores

```
modelo_beta <- function(params, n = 5000){
  rbeta(n, params$alpha, params$beta)
}
```

```r
escenarios <-
  tibble(analista = fct_inorder(c("Ignorante", "Indiferente",
                                  "Feminista", "Ingenuo")),
         alpha = c(1, .5, 5, 14),
         beta  = c(1, .5, 11, 1)) |>
  nest(params.previa = c(alpha, beta)) |>
  mutate(muestras.previa = map(params.previa, modelo_beta))
```

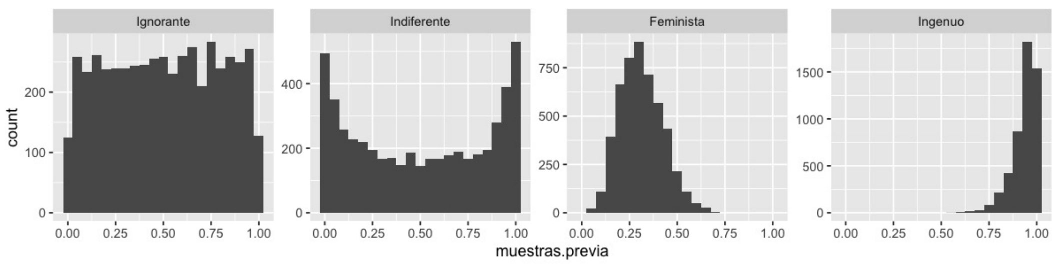


FIGURA 2. *Muestras de* $\theta \sim$ Previa .

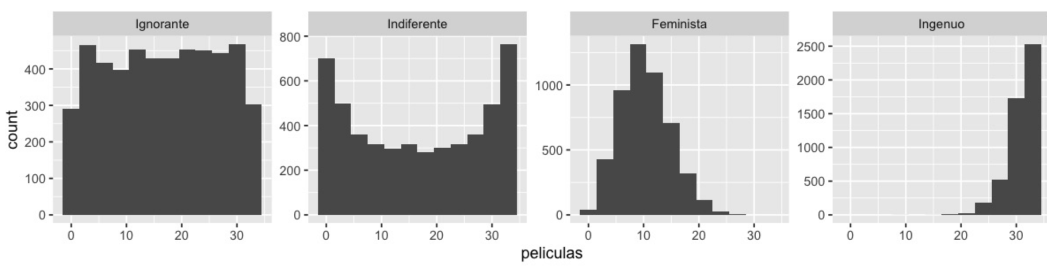
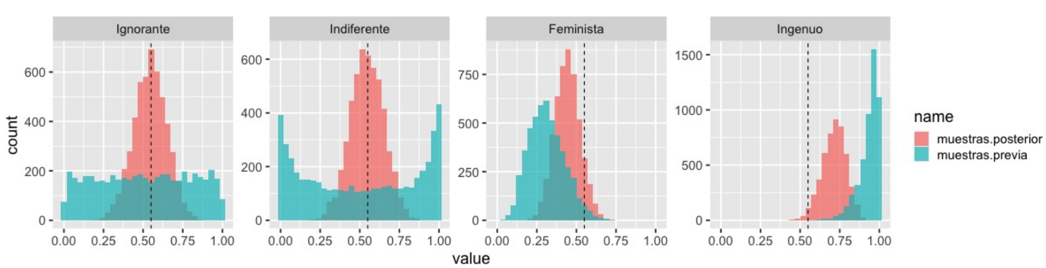


FIGURA 3. *Distribución predictiva previa*

```r
update_rule <- function(params){
  tibble(alpha = params$alpha + data$PASS,
         beta  = params$beta  + data$FAIL)
}
escenarios <- escenarios |>
  mutate(params.posterior = map(params.previa, update_rule),
         muestras.posterior = map(params.posterior, modelo_beta))
```



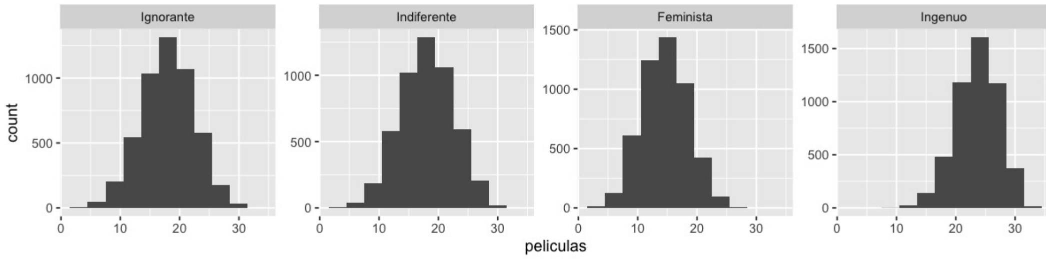
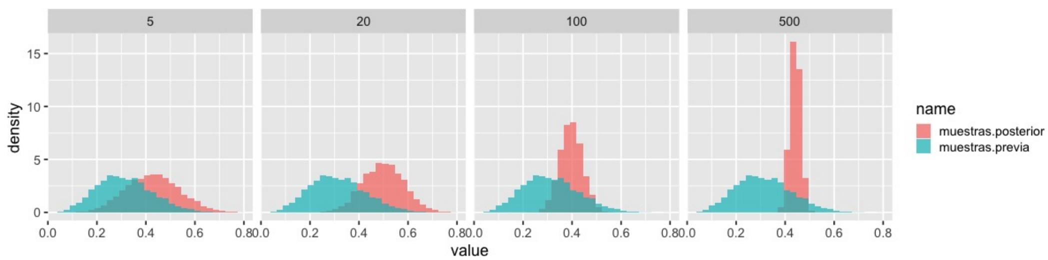


Figura 4. *Predictiva posterior.*

3.3. Diferentes datos, diferentes posteriores



4. MOTIVACIÓN

Por medio de metodología Bayesiana podemos cuantificar incertidumbre en:

- Observaciones.
- Parámetros.
- Estructura.

Es fácil especificar y ajustar modelos. Pero hay preguntas cuyas respuestas no han quedado claras:

1. Construcción.
2. Evaluación.
3. Uso.

> Programación probabilística.

Los aspectos del flujo de trabajo Bayesiano consideran ([2]):

1. Construcción iterativa de modelos.
2. Validación de modelo (computacional).
3. Entendimiento de modelo.
4. Evaluación de modelo.

4.1. Distinción importante

`Inferencia` no es lo mismo que `análisis de datos` o que un `flujo de trabajo`.

> Inferencia (en el contexto bayesiano) es formular y calcular con probabilidades condicionales.

4.2. ¿Por qué necesitamos un flujo de trabajo?

- El cómputo puede ser complejo.

- Expandir nuestro entendimiento en aplicaciones.
- Entender la relación entre modelos.
- Distintos modelos pueden llegar a distintas conclusiones.

4.3. Proceso iterativo

- La gente de ML sabe que el proceso de construcción de un modelo es iterativo, ¿por qué no utilizarlo?

> Una posible explicación puede encontrarse en [1]. El argumento es formal en cuanto a actualizar nuestras creencias como bayesianos. Sin embargo, con cuidado y un procedimiento científico puede resolver el asunto.

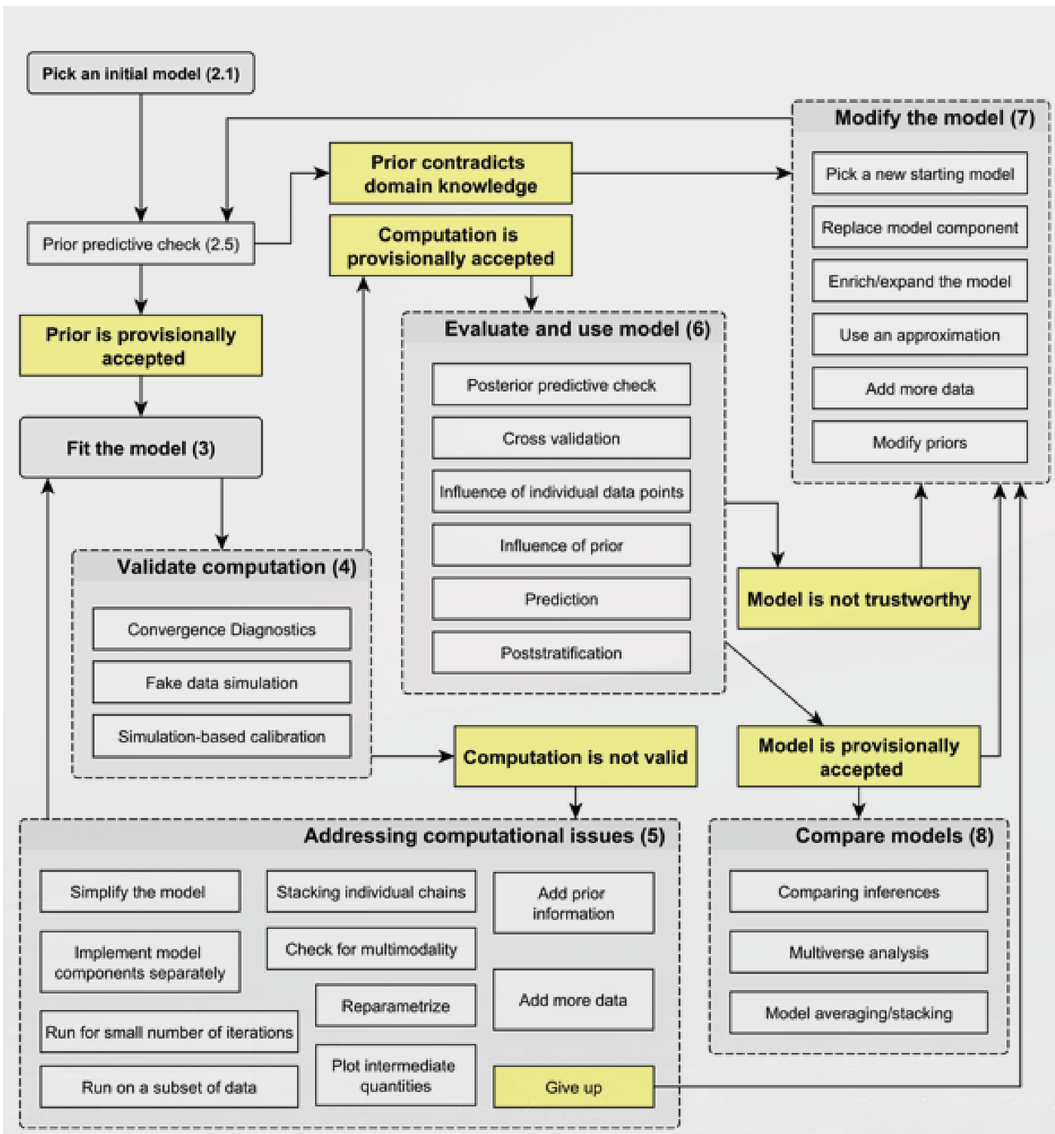


FIGURA 5. *Tomado de [2].*

REFERENCIAS

[1] A. Gelman and Y. Yao. Holes in Bayesian statistics. *Journal of Physics G: Nuclear and Particle Physics*, 48(1):014002, jan 2021. ISSN 0954-3899, 1361-6471. . 5

[2] A. Gelman, A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák. Bayesian workflow. *arXiv preprint arXiv:2011.01808*, 2020. 1, 4, 5

[3] A. Johnson, M. Ott, and M. Dogucu. *Bayes Rules! An Introduction to Applied Bayesian Modeling.* 2021. 2

[4] J. Kruschke. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Academic Press, 2014. 2