

EST-46115: Modelación Bayesiana

Profesor: Alfredo Garbuno Iñigo — Primavera, 2023.

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 recomendadas: Notas del [curso de fundamentos](#) (2021) y sección 1 de [2].

1. MOTIVACIÓN

2. NOTACIÓN

Denotamos por x una **variable aleatoria** y por $\mathbb{P}(\cdot)$ una **función de distribución**. Escribimos $x \sim \mathbb{P}$ para denotar que la variable aleatoria x tiene distribución $\mathbb{P}(\cdot)$. Denotamos por $\mathbb{E}[\cdot]$ el **valor esperado** del argumento con respecto a la distribución que estamos considerando. Durante el curso seremos explícitos en la variable aleatoria y usaremos

$$\mathbb{E}_x[\cdot] = \int_{\mathcal{X}} \cdot \pi(x) dx, \quad (1)$$

o bien, haremos énfasis en la distribución por medio de lo siguiente

$$\mathbb{E}_{\pi}[\cdot] = \int_{\mathcal{X}} \cdot \pi(x) dx, \quad (2)$$

de acuerdo al contexto.

Nota que en las ecuaciones anteriores estamos considerando el término $\pi(\cdot)$ como la **función de densidad** de la **función de probabilidad** $\mathbb{P}(\cdot)$.

Nos será útil la siguiente notación para evaluar valores esperados

$$\pi(f) := \mathbb{E}_{\pi}[f(x)] = \int_{\mathcal{X}} f(x) \pi(x) dx, \quad (3)$$

pues será el **objetivo general** para los métodos que estudiaremos en el curso.

Por ejemplo, utilizaremos la noción de **aproximar integrales** por medio de algún procedimiento de muestreo de tal forma que esperaremos construir una estimación $\hat{\pi}(f)$ con cierto grado de refinamiento. Por ejemplo, veremos el **método Monte Carlo** que utiliza una colección de N simulaciones para aproximar la integral anterior. Esto lo denotaremos por

$$\hat{\pi}_N(f) \approx \pi(f). \quad (4)$$

En general, nos interesa, y esperamos que, podamos:

1. Mejorar nuestra estimación con mas simulaciones

$$\lim_{N \rightarrow \infty} \hat{\pi}_N(f) = \pi(f) \quad (5)$$

2. Cuantificar la incertidumbre en nuestra aproximación por medio de alguna distribución de probabilidad. Por ejemplo,

$$\hat{\pi}_N(f) \sim \mathcal{N}\left(\pi(f), \frac{\mathbb{V}(f)}{N}\right). \quad (6)$$

2.0.1. Definición [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

$$\pi_\theta(x) \quad \text{ó} \quad \pi(x; \theta), \quad (7)$$

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

3. REPASO DE PROBABILIDAD

Consideraremos como requisitos el contenido de **Cálculo de Probabilidades II** y **Álgebra Lineal** (o equivalentes). En particular lo que requerimos como base es lo siguiente.

3.0.1. Definición [Espacio de Probabilidad]: Un espacio de probabilidad está definido por la terna $(\Omega, \mathcal{X}, \mathbb{P})$:

1. El espacio muestral, Ω (elementos).
2. El espacio de eventos medibles, \mathcal{X} (subconjuntos).
3. La medida de probabilidad, $\mathbb{P} : \mathcal{X} \rightarrow [0, 1]$.

3.0.2. Definición [Variable aleatoria]: Una variable aleatoria es una función $X : \mathcal{X} \rightarrow \mathbb{R}$ con la propiedad de que las pre-ímagenes bajo X son eventos medibles. Es decir,

$$\{w \in \mathcal{X} : X(w) \leq x\} \in \mathcal{X} \quad \forall x \in \mathbb{R}. \quad (8)$$

3.0.3. Definición [Función de acumulación]: Para toda variable aleatoria X tenemos una función de acumulación $\mathbb{P}_X : \mathbb{R} \rightarrow [0, 1]$ dada por

$$\mathbb{P}_X(x) = \mathbb{P}(\{w \in \mathcal{X} : X(w) \leq x\}). \quad (9)$$

Esto usualmente lo escribimos como $\mathbb{P}_X(x) = \mathbb{P}\{X \leq x\}$.

3.0.4. Definición [Función de densidad]: Una variable aleatoria es continua si su función de acumulación es **absolutamente continua** y puede ser expresada por medio de

$$\mathbb{P}_X(x) = \int_{-\infty}^x \pi(s) ds, \quad (10)$$

donde la anti-derivada $\pi : \mathbb{R} \rightarrow [0, \infty)$ se llama la **función de densidad** de la variable aleatoria X .

Las propiedades generales de las distribuciones de probabilidad se pueden especificar por medio de su centralidad (localización), su dispersión, su rango de valores, su simetría y el comportamiento de valores extremos.

En general esto lo podemos extraer de los momentos

$$\mathbb{E}(X^p) = \int_{\mathbb{R}} x^p \pi(x) dx, \quad (11)$$

o los momentos centrales. Por ejemplo: media y varianza.

Uno de los resultados que espero recuerden bien de sus cursos anteriores es el de la **Ley de los Grandes Números**. La cual podemos enunciar como:

3.0.5. Teorema [Ley de los Grandes Números]: Sea X_1, X_2, \dots una colección de variables aleatorias independientes e idénticamente distribuidas (iid) y sea \bar{X}_n el promedio de un subconjunto de n . Si denotamos por μ el valor promedio de X_i dentro de esa colección, entonces tenemos que

$$\bar{X}_n \rightarrow \mu \quad (\text{casi seguramente}). \quad (12)$$

3.0.6. **Teorema [Límite Central]:** Sea X_1, \dots, X_n una colección de n variables aleatorias iid con $\mathbb{E}[X_i] = \mu$ y $\mathbb{V}[X_i] = \sigma^2 < \infty$. Entonces

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad (13)$$

para n suficientemente grande.

4. REPASO INFERENCIA

Repaso de inferencia bajo un enfoque frecuentista.

4.1. Regla de Bayes

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

$$\pi(A|B) = \frac{\pi(B|A)\pi(A)}{\pi(B)}. \quad (14)$$

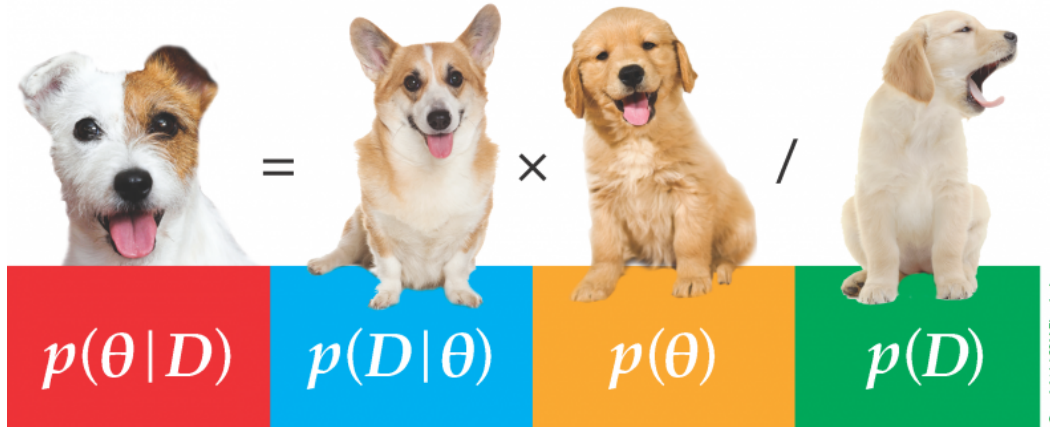


FIGURA 1. Tomado de [4].

4.2. Ejemplos

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

5. REPASO INFERENCIA

Repaso de inferencia bajo un enfoque bayesiano.

5.1. Ejemplo

Este ejemplo fue tomado de [3].

5.2. Diferentes previas, diferentes posteriores

```

1 modelo_beta <- function(params, n = 5000){
2   rbeta(n, params$alpha, params$beta)
3 }

```

```

1 escenarios <-
2   tibble(analista = fct_inorder(c("Ignorante", "Indiferente",
3     "Feminista", "Ingenuo")),
4     alpha = c(1, .5, 5, 14),
5     beta = c(1, .5, 11, 1)) >
6   nest(params.previa = c(alpha, beta)) >
7   mutate(muestras.previa = map(params.previa, modelo_beta))

```

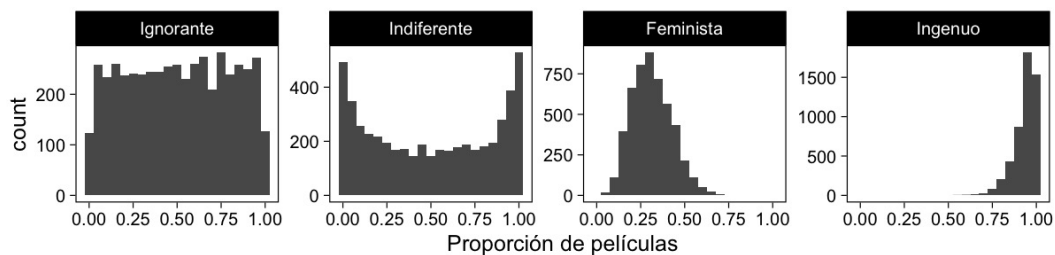
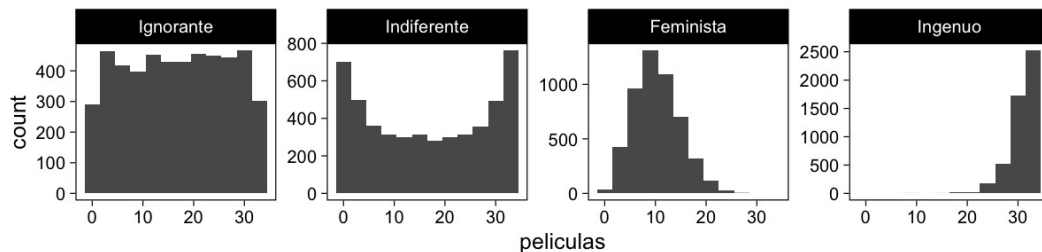
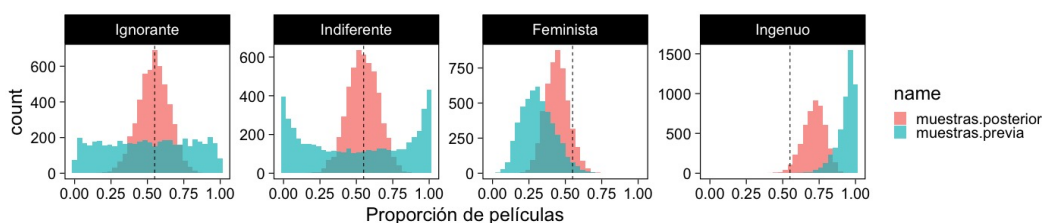
FIGURA 2. Muestras de $\theta \sim \text{Previa}$.

FIGURA 3. Distribución predictiva previa

```

1 update_rule <- function(params){
2   tibble(alpha = params$alpha + data$PASS,
3     beta = params$beta + data$FAIL)
4 }
5 escenarios <- escenarios >
6   mutate(params.posterior = map(params.previa, update_rule),
7     muestras.posterior = map(params.posterior, modelo_beta))

```



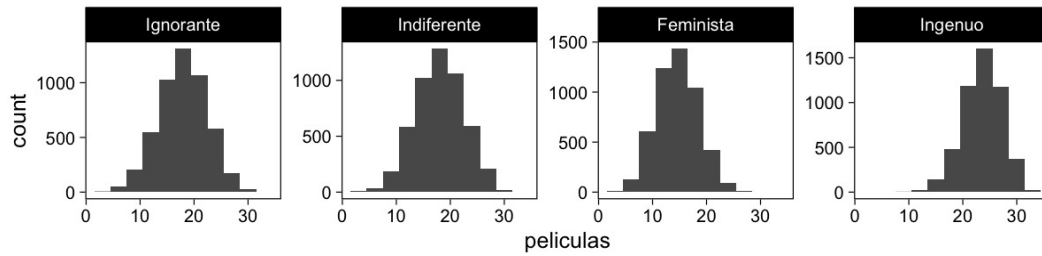
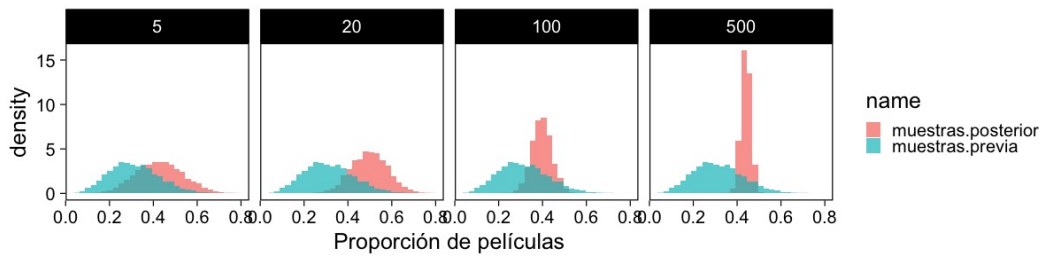


FIGURA 4. Predictiva posterior.

5.3. Diferentes datos, diferentes posteriores



5.4. Análisis secuencial

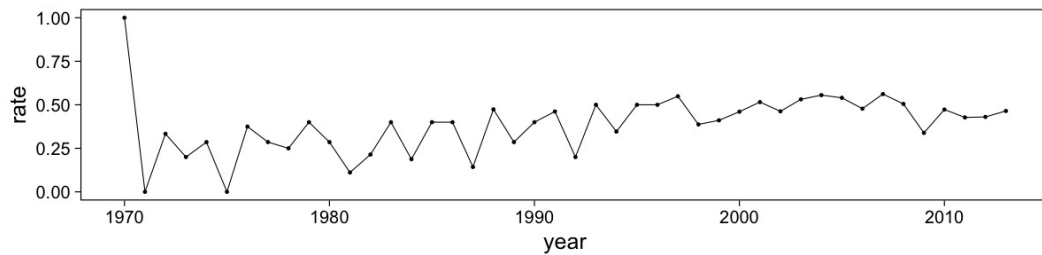


FIGURA 5. Histórico de la proporción de películas que pasan la prueba de Bechdel por año.

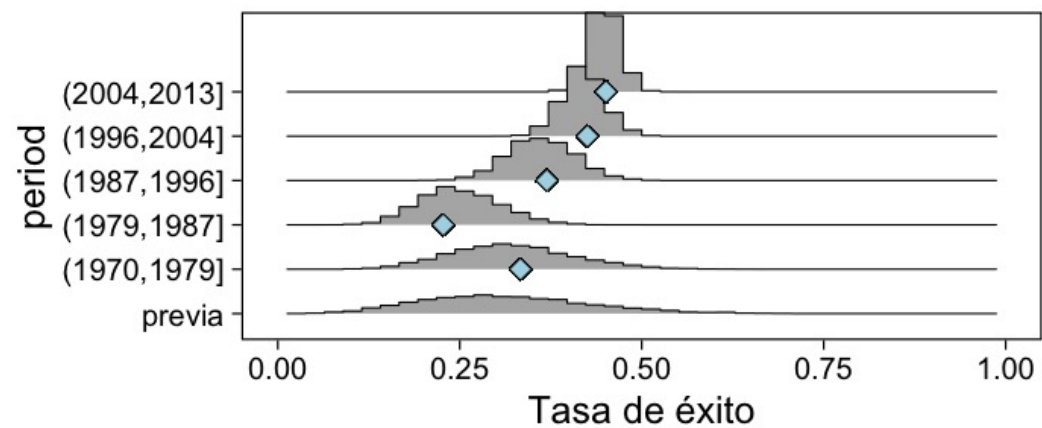


FIGURA 6. La posterior de hoy puede ser la previa de mañana.

5.5. Tarea

Echenle un ojo a la sección 5.2 de [Bayes rules!](#) donde se expone a detalle un modelo más del análisis conjugado. ¿Puedes identificar/derivar la distribución predictiva?

6. 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.

6.1. 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.

6.2. ¿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.

6.3. 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.

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. . 6

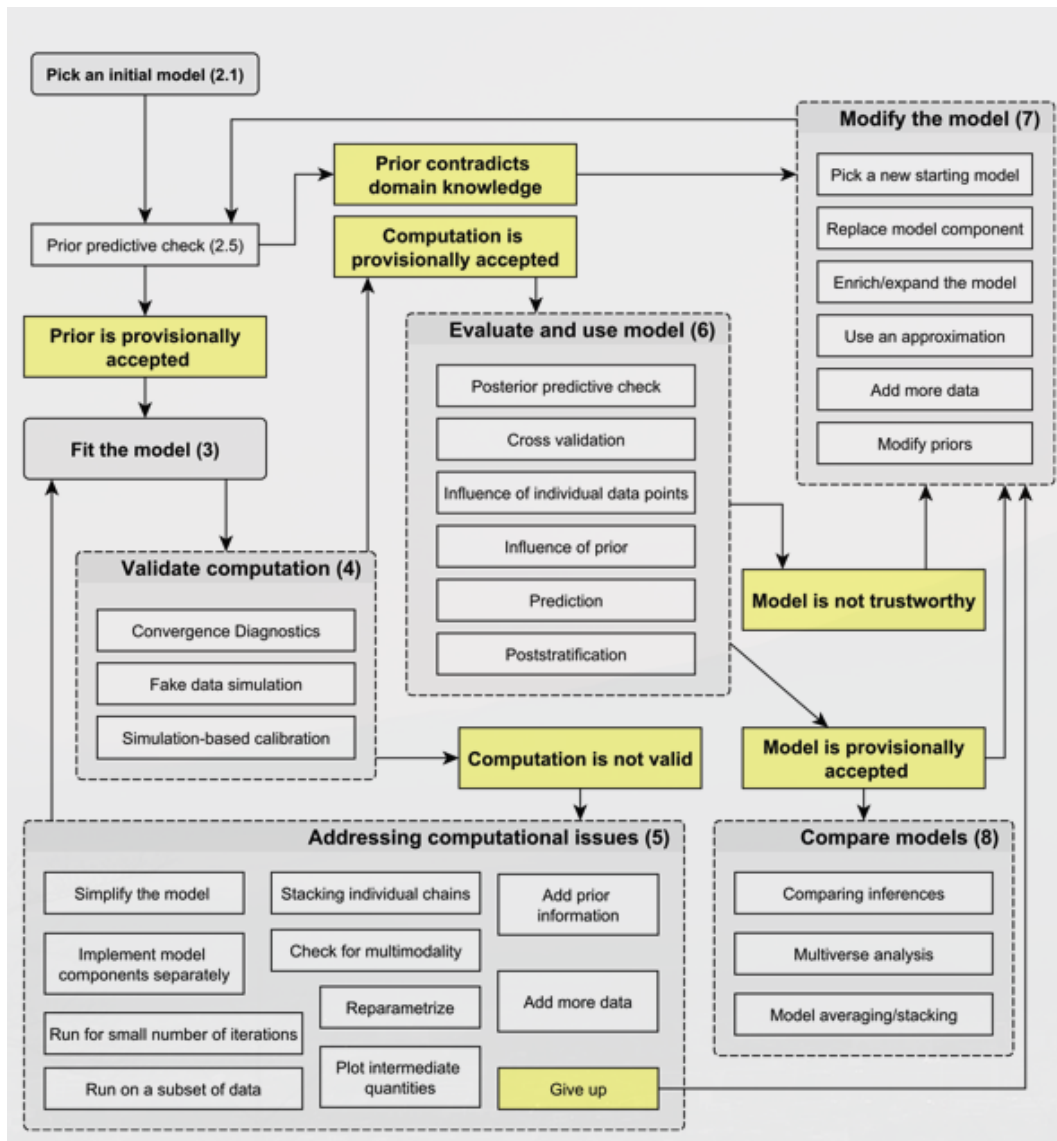


FIGURA 7. Tomado de [2].

- [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, 6, 7
- [3] A. Johnson, M. Ott, and M. Dogucu. *Bayes Rules! An Introduction to Applied Bayesian Modeling*. 2021. 3
- [4] J. Kruschke. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press, 2014. 3