

EST-46115: Modelación Bayesiana

Profesor: Alfredo Garbuno Iñigo — Primavera, 2022 — Toma de decisiones.

Objetivo: En esta sección conectaremos conceptos que hemos visto durante el curso para toma de decisiones bajo incertidumbre. Anteriormente, introdujimos el concepto de inferencia Bayesiana dentro del marco de teoría de la decisión. Retomaremos ciertos componentes de este marco en el contexto de inferencia y modelado predictivo..

Lectura recomendada: Sección 9.9 de [3]. Capítulo 9 de [1]. Algunos pasajes de [4]. el libro de Gilboa [2] es una excelente referencia para el modelado de toma de decisiones bajo incertidumbre.

1. INTRODUCCIÓN

Retomaremos la discusión sobre toma de decisiones bajo incertidumbre. Para tomar decisiones debemos de especificar la noción de **utilidad** asociada a cada opción que podemos tomar. La decisión **óptima Bayesiana** corresponde a la que maximiza la utilidad esperada. Buscaremos ejemplificar cómo podemos utilizar **Stan** para estimar la distribución de posibles respuestas bajo decisiones y calcular utilidades esperadas.

2. DEFINICIÓN DE UN PROBLEMA DE DECISIÓN

Siguiendo a [1] un problema de decisión Bayesiano necesita de los siguientes componentes:

1. Definir un conjunto de posibles resultados X .
2. Definir un conjunto de posibles decisiones D .
3. Definir una función de utilidad que contraste la decisión, d , contra el resultado x , la cual denotamos por $U(x, d)$.
4. Especificar nuestro estado de conocimiento sobre las posibles realizaciones de posibles resultados a través de $\pi(x)$.

También es usual, como veremos adelante, que consideremos una $\pi(x)$ para cada decisión que podamos tomar, lo cual lo denotamos por $\pi(x|d)$ si fuera necesario.

Con lo cual podemos escoger la decisión d^* que obtenga la mejor utilidad esperada

$$d^* = \arg \max_d \bar{U}[d], \quad (1)$$

donde

$$\bar{U}[d] = \mathbb{E}[U(X, d)] = \int U(x, d)\pi(x)dx. \quad (2)$$

Los resultados deberán de representar la mayor información posible que sea relevante para la especificación de la función de utilidad.

2.1. Contexto bayesiano

- Las decisiones pueden ser sobre los parámetros del modelo, θ , o cantidades observables, \tilde{y} .
- Nuestro estado de conocimiento lo definimos como la distribución posterior o predictiva posterior.
- La función de utilidad depende del contexto del problema.

2.2. Caso: enfoque predictivo

Supongamos que tenemos el siguiente problema de decisión bayesiano con un **enfoque predictivo**:

- Los **estados** posibles son **cantidades observables** \tilde{y} .
- Las **decisiones** que podemos tomar son sobre *todas* las posibles **funciones de probabilidad** relevantes para nuestro problema. Es decir, cualquier $d(\tilde{y})$ donde d es una distribución de probabilidad sobre cantidades observables.
- La **función de utilidad** que escogeremos será **utilidad logarítmica**, $\log(d)$.
- Nuestro **estado de conocimiento** sobre los estados inciertos lo reflejamos a través de la **distribución predictiva posterior**, $\pi(\tilde{y}|\underline{y}_n)$.

La decisión que maximiza la utilidad esperada bajo nuestro estado de conocimiento será la que **maximice**

$$\int \log d(\tilde{y}) \pi(\tilde{y}|\underline{y}_n) d\tilde{y}. \quad (3)$$

Nota que pedimos que sea la que **minimice**

$$- \int \log d(\tilde{y}) \pi(\tilde{y}|\underline{y}_n) d\tilde{y}, \quad (4)$$

que es justamente la **entropía cruzada** entre dos distribuciones y que sabemos tiene un punto óptimo siempre y cuando utilicemos la misma distribución con la que reflejamos nuestro estado de conocimiento.

Es decir, bajo un **enfoque predictivo** nuestra mejor decisión bajo utilidad logarítmica es utilizar la densidad predictiva posterior.

2.3. Caso: enfoque de inferencia

Supongamos que tenemos el siguiente problema de decisión bayesiano con un **enfoque de inferencia**:

- Los **estados** posibles son la **configuración del modelo** que especifica un modelo probabilístico, θ .
- Las **decisiones** que podemos tomar son *todas* las posibles **funciones de probabilidad** relevantes para nuestro problema. Es decir, cualquier $d(\theta)$ donde d es una distribución sobre configuraciones de un modelo.
- La **función de utilidad** que escogeremos será **utilidad logarítmica**, $\log(d)$.
- Nuestro **estado de conocimiento** sobre los estados inciertos lo reflejamos a través de la **distribución posterior**, $\pi(\theta|\underline{y}_n)$.

2.3.1. *Pregunta:* ¿Cuál será la mejor decisión que podemos tomar en este escenario?

3. ANÁLISIS DE DECISIÓN

Vamos a seguir el [ejemplo](#) que está en la documentación de **Stan**. En este escenario el tomador de decisiones tiene que decidir el medio de transporte para llegar a su trabajo: caminar, bicicleta, transporte público o taxi.

A lo largo del año ha registrado 200 días de trayectos a su trabajo y ha registrado el tiempo que le toma llegar.

3.1. Definición de decisiones y observaciones

- Las **decisiones** son el medio de transporte codificadas numéricamente.
- Los **resultados** $X = \mathbb{R} \times \mathbb{R}$ que observamos son el tiempo t que toma y el costo c asociado a ese tiempo, $x = (c, t)$.

3.2. Definición de estado de conocimiento

Necesitamos definir $\pi(x|d)$ la distribución de resultados posibles sujeta a la decisión que se ha tomado. Bajo el enfoque Bayesiano ésta será la distribución predictiva posterior de una observación condicional en la historia que hemos visto

$$\pi(\tilde{x}|d, \underline{x}_n, \underline{d}_n) = \int \pi(\tilde{x}|d, \theta) \pi(\theta|\underline{x}_n, \underline{d}_n) d\theta. \quad (5)$$

Por simplicidad utilizamos una distribución log-normal para los tiempos de llegada bajo cada transporte. Es decir, para una observación $x_n = (c_n, t_n)$ asociada a la decisión d_n consideramos

$$t_n \sim \text{LogNormal}(\mu_{[d_n]}, \sigma_{[d_n]}) . \quad (6)$$

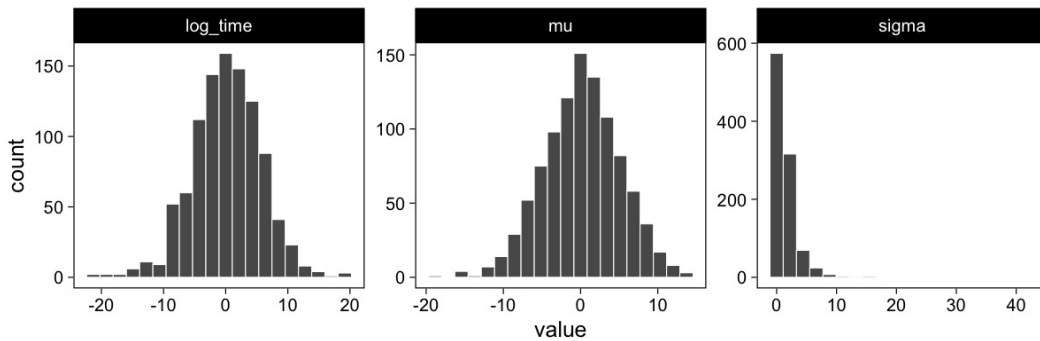
$$c_n \sim \text{LogNormal}(\nu_{[d_n]}, \tau_{[d_n]}) . \quad (7)$$

Decimos que una variable aleatoria se distribuye log-normal, denotado como $Y \sim \text{logNormal}(\mu, \sigma)$, si $\log Y \sim \text{Normal}(\mu, \sigma)$.

Las previas que utilizamos para el tiempo de llegada en cada modo de transporte, $k \in \{1, \dots, 4\}$, son

$$\mu_k \sim \text{Normal}(0, 5) , \quad (8)$$

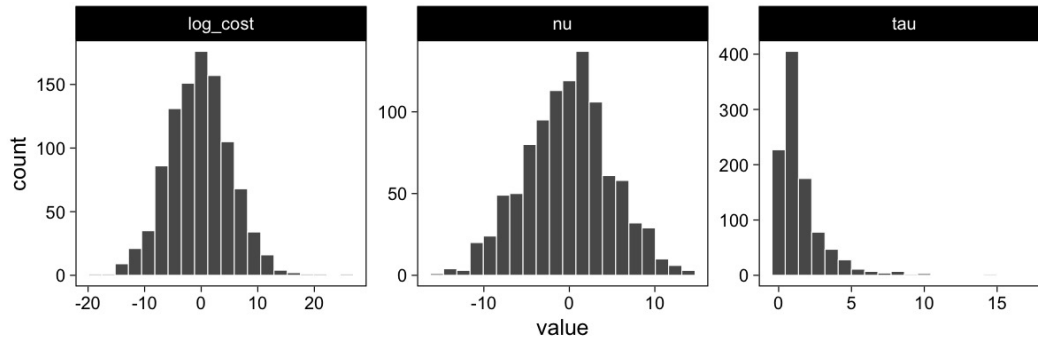
$$\sigma_k \sim \text{logNormal}(0, 1) . \quad (9)$$



Las previas que utilizamos para los costos por cada modo de transporte, $k \in \{1, \dots, 4\}$, son

$$\nu_k \sim \text{Normal}(0, 5) , \quad (10)$$

$$\tau_k \sim \text{logNormal}(0, 1) . \quad (11)$$



El conjunto de parámetros del modelo que marginalizará en la predictiva posterior es

$$\theta = (\mu_{1:4}, \sigma_{1:4}, \nu_{1:4}, \tau_{1:4}). \quad (12)$$

3.3. Definición función de utilidad

Digamos que el tomador de decisión evalúa su tiempo de traslado de manera lineal y que el tiempo invertido en transporte lo evalúa en \$25 por cada momento que éste pasa en su trayecto, por lo que la función de utilidad es

$$U(c, t) = -(c + 25 \cdot t). \quad (13)$$

Nota que podríamos considerar una utilidad distinta para cada modo de transporte, $U(x, d)$, de tal manera que se reflejen costos individuales de cada medio de transporte.

3.4. Cálculo de utilidad esperada

Lo que necesitamos ahora es poder calcular la utilidad esperada de cada una de las posibles decisiones y tomar la que minimice dicha función. El siguiente código aprovecha que nuestro espacio de posibles decisiones es pequeño.

```

1 functions {
2   real U(real c, real t) {
3     return -(c + 25 * t);
4   }
5 }
6 data {
7   int<lower=0> N;
8   array[N] int<lower=1, upper=4> d;
9   array[N] real c;
10  array[N] real<lower=0> t;
11 }
12 parameters {
13   vector[4] mu;
14   vector<lower=0>[4] sigma;
15   array[4] real nu;
16   array[4] real<lower=0> tau;
17 }
18 model {
19   mu ~ normal(0, 1);
20   sigma ~ lognormal(0, 0.25);
21   nu ~ normal(0, 20);
22   tau ~ lognormal(0, 0.25);
23   t ~ lognormal(mu[d], sigma[d]);

```

```
24   c ~ lognormal(nu[d], tau[d]);
25 }
26 generated quantities {
27   array[4] real util;
28   for (k in 1:4) {
29     util[k] = U(lognormal_rng(mu[k], sigma[k]),
30                 lognormal_rng(nu[k], tau[k]));
31   }
32 }
```

Lo que esta calculando Stan son los términos para estimar la utilidad esperada por medio de un estimador Monte Carlo. Esto lo vemos de la expresión

$$\bar{U}[d] = \mathbb{E}[U(X, d)|\underline{x}_n, \underline{d}_n] = \int U(x, d) \cdot \pi(x|d, \theta) \cdot \pi(\theta|\underline{x}_n, \underline{d}_n) d\theta dx, \quad (14)$$

$$\approx \frac{1}{M} \sum_{m=1}^M U(x^{(m)}), \quad (15)$$

donde

$$x^{(m)} \sim \pi(x|d, \theta^{(m)}), \quad (16)$$

$$\theta^{(m)} \sim \pi(\theta|\underline{x}_n, \underline{d}_n). \quad (17)$$

4. FUNCIONES DE UTILIDAD

```
1 functions {
2   real quadraticUtility(int y_tilde, int d) {
3     return -(y_tilde - d)^2;
4   }
5   real zeroOneUtility(int y_tilde, int d){
6     if (y_tilde == d) {
7       return 1;
8     } else {
9       return 0;
10    }
11  }
12  real intervalUtility(int y_tilde, int d){
13    if (fabs(y_tilde - d) < 10) {
14      return 0;
15    } else {
16      return -exp(-fabs(y_tilde - d));
17    }
18  }
19 }
20
21 data {
22   int<lower=0> N;
23   int<lower=0> K;
24   array[N] int<lower=0, upper=1> test;
25 }
26
27 parameters {
28   real<lower=0, upper=1> theta;
29 }
30
31 model {
```

```

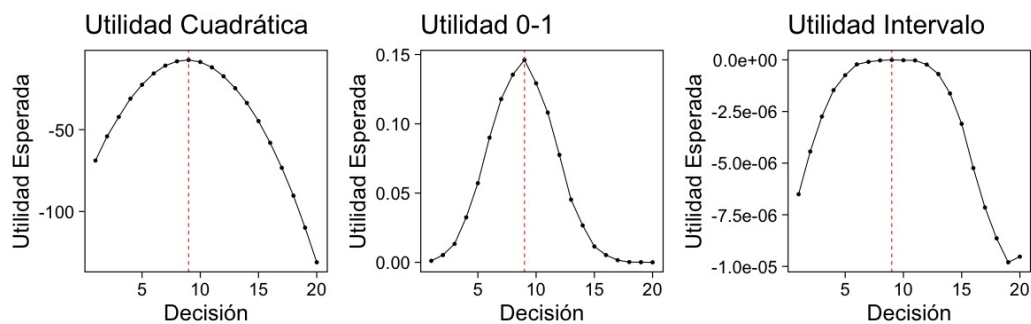
32  theta ~ beta(5, 11);
33  test ~ bernoulli(theta);
34  }
35
36  generated quantities {
37    array[K] real utilQuad;
38    array[K] real utilZeroOne;
39    array[K] real utilInterval;
40    for (kk in 1:K) {
41      utilQuad[kk] = quadraticUtility(binomial_rng(K, theta), kk);
42      utilZeroOne[kk] = zeroOneUtility(binomial_rng(K, theta), kk);
43      utilInterval[kk] = intervalUtility(binomial_rng(K, theta), kk);
44    }
45  }

```

```

1  posterior <- modelo$sample(data = c(N = nrow(data),
2                                     K = 20,
3                                     data),
4                                refresh = 4000,
5                                iter_sampling = 4000,
6                                seed = 108727)

```



5. DECISIONES CONTINUAS

El ejemplo anterior utilizaba decisiones discretas (o un espacio de decisiones discretas). Si las decisiones fueran sobre un continuo el problema se vuelve mas complicado, para lo cual las capacidades actuales de **Stan** son limitadas.

REFERENCIAS

- [1] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*, volume 2. CRC press Boca Raton, FL, 2014. [1](#)
- [2] I. Gilboa. *Theory of Decision under Uncertainty*. [1](#)
- [3] O. A. Martin, R. Kumar, and J. Lao. *Bayesian Modeling and Computation in Python*. Chapman and Hall/CRC, Boca Raton, First edition, 2021. [1](#)
- [4] A. Vehtari and J. Ojanen. A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys*, 6(none):142–228, jan 2012. ISSN 1935-7516. . [1](#)