9 Teoría de la Decisión

En condiciones de incertidumbre, el papel del estadístico es tomar una decisión a partir de una observación parcial de la situación. Una mala decisión puede llevar a perdidas importantes. Para comparar las decisiones definimos una función de perdidas que mide las consecuencias de cada decisión.
Consideramos el ejemplo siguiente: Un estadístico tiene que optar cada mañana entre dos trayectos. La duración de cada uno depende del estado del trafico que para simplificar clasificaremos en fluido (10% de las veces), normal (60% de las veces) y malo (30% de los casos). Según el estado del tráfico se obtienen los tiempos de trayecto :

Suceso	Probabilidad Q	trayecto d_1	trayecto d_2
$\theta_1 = F$	0.1	$15m$	$30m$
$\theta_2 = N$	0.6	$35m$	$40m$
$\theta_3 = M$	0.3	$70m$	$50m$

El riesgo de una decisión es aquí debido al hecho de que no podemos conocer de manera segura el estado del trafico en un momento dado. Para comparar las decisiones (los trayectos posibles) utilizamos aquí la duración media del trayecto:

$$\overline{w}_Q(d_i) = d_i(\theta_1)Q(\theta_1) + d_i(\theta_2)Q(\theta_2) + d_i(\theta_3)Q(\theta_3)$$

Obtenemos que $\overline{w}_Q(d_1) = 43.5m$ y $\overline{w}_Q(d_2) = 42m$ y concluimos que el trayecto d_2 (o decisión d_2) es, en promedio, mejor que d_1.

El marco general es el siguiente:

1. Un conjunto Θ cuyos elementos θ describen los varios estados de la situación.

2. Un conocimiento *a priori* sobre la distribución Q de los estados de la situación.

3. El estadístico puede realizar un experimento cuyo resultado X depende de θ. Este resultado, le permitirá inferir sobre el valor de θ.

4. Luego, definimos un conjunto D de decisiones que el estadístico podrá tomar. En los problemas de estimación el conjunto D coincide con el

conjunto Θ (tomar una decisión es seleccionar un valor en Θ). Para los problemas de contrastes, D es el conjunto de las hipótesis $D = \{H_0, H_1\}$.

5. La función de perdidas $w(d, \theta)$ se define en el espacio $D \times \Theta$ y mide las perdidas si se toma la decisión d cuando la situación estudiada está en el estado θ.

Si omitimos el punto (3) asociado a una observación de la situación, obtenemos un *juego con dos jugadores* (ejemplo: el estadístico frente a la naturaleza)

9.1 Juego con dos jugadores

Definicin 7 *Un juego con dos jugadores se define por el conjunto D de decisiones d (o estrategias) del primer jugador, los elementos $\theta \in \Theta$, estrategias del segundo jugador y una función de perdidas $w(d, \theta)$ positiva, que mide las perdidas sufridas por el primer jugador si elige la estrategia d cuando el segundo elige θ.*

El problema fundamental de la teoría de los juegos consiste en hallar la estrategia optima del primer jugador con el cual nos identificamos.

Definicin 8 *Una estrategia d_1 es mejor que d_2 si para cualquier $\theta \in \Theta$,*

$$w(d_2, \theta) \geq w(d_1, \theta),$$

y $w(d_2, \theta) > w(d_1, \theta)$ para al menos un θ.

9.1.1 Estrategia Bayesiana

Este tipo de estrategia se utiliza cuando el segundo jugador elige su decisión (ejemplo: la naturaleza) con una cierta distribución Q (*a priori*) definida sobre el conjunto Θ. Para comparar las estrategias $d \in D$, utilizamos en este caso la esperanza de la función de perdidas llamado riesgo bayesiano:

$$\overline{w}_Q(d) = E_Q\left(w(d, \theta)\right) = \int w(d, \theta) q(\theta) d\theta$$

done $q(\theta)$ es la función de densidad de Q.

Definicin 9 *Llamamos estrategia bayesiana asociada a una distribución a priori Q, la estrategia d_Q que minimiza el riesgo bayesiano $\overline{w}_Q(d)$:*

$$\overline{w}_Q(d_Q) = \min_{d \in D} \overline{w}_Q(d).$$

Ejemplo 16 *Un ingeniero tiene que construir un dique contra las inundaciones debidas a las crecidas de un río. Si decide construir un dique con altura d (a partir del nivel medio del río) costará cd Millones de pesetas ; si hay una crecida, habrá ningún daño si su altura θ (a partir del nivel medio) es inferior a d. Pero, una perdidas cifradas a $C(\theta - d)$ Millones de pesetas ($C > c$) si $\theta > d$. La altura de una crecida θ del río es una variable que suponemos exponencial $Q =Exp(\lambda)$ con parámetro λ. La función de perdidas asociada a la decisión d será entonces:*

$$w(d, \theta) = cd + C(\theta - d)I_{(\theta > d)}.$$

Por tanto, el riesgo bayesiano es

$$\overline{w}_Q(d) = cd + \frac{C}{\lambda} \exp(-d\lambda).$$

Puesto que

$$\begin{aligned} E_\lambda \left[(\theta - d) I_{\{\theta > d\}} \right] &= \int_d^\infty (x - d)\lambda \exp(-\lambda x) dx \\ &= \int_0^\infty u\lambda \exp(-\lambda(u + d)) dx \\ &= \exp(-\lambda d) \int_0^\infty u\lambda \exp(-\lambda u) dx \\ &= \exp(-\lambda d)/\lambda \end{aligned}$$

Si λ es conocido, el ingeniero tendrá que optar por la altura d_Q que minimiza $\overline{w}_Q(d)$. Tenemos que

$$\frac{\partial \overline{w}_Q(d)}{\partial d} = c - C \exp(-d\lambda) = 0$$

si $d_Q = \frac{1}{\lambda} \ln(C/c)$. La función de perdida es para la altura d_Q igual a

$$\begin{aligned} \overline{w}_Q(d_Q) &= c\frac{1}{\lambda} \ln(C/c) + \frac{C}{\lambda} \exp\left(-\frac{1}{\lambda} \ln(C/c)\lambda \right) \\ &= \frac{c}{\lambda} \left[\ln(C/c) + 1 \right]. \end{aligned}$$

Este valor corresponde al riesgo mínimo para un conocimiento *a priori* de la situación.

Teorema 10 *Si la función de perdidas es cuadrática:* $w(d,\theta) = (d-\theta)^2$ *entonces, la estrategía bayesiana asociada a una distribución a priori* Q *es la media de* $\theta : d_Q = E_Q(\theta)$*. Además, tenemos que* $\overline{w}_Q(d_Q) = var_Q(\theta)$*.*

Prueba : Sea d una decisión, por el teorema de Pitágoras se puede descomponer el riesgo bayesiano $\overline{w}_Q(d)$ de la siguiente manera:

$$\begin{aligned}
\overline{w}_Q(d_Q) &= E_Q\left[(d-\theta)^2\right] \\
&= E_Q\left[(d-E_Q(\theta))^2\right] + 2E_Q\left[(d-E_Q(\theta))\left(E_Q(\theta)-\theta\right)\right] \\
&\quad + E_Q\left[(E_Q(\theta)-\theta)^2\right] \\
&= [d-E_Q(\theta)]^2 + \mathrm{var}_Q(\theta)
\end{aligned}$$

Ejemplo 17 *Predicción del número* θ *de peces en un lago: Por experiencia de los años anteriores, se sabe (* a priori) *que este número sigue una distribución* Q *normal* $N(500, 40)$*. Si elegimos una función de perdidas cuadrática* $w(d,\theta) = (d-\theta)^2$*, la mejor decisión (aquí mejor predicción) con este nivel de conocimiento de* θ*, es* $d_Q = E_Q(\theta) = 500$*, y la perdida o error mínimo será* $\overline{w}_Q(d_Q) = var(\theta) = 40$*.*

9.2 Decisión con información muestral

Se trata aquí de una decisión basada en una observación de la situación. Supongamos que dispongamos de una observación X cuya distribución $p(x|\theta)$ depende de θ (variable aleatoría que describe la situación). En este caso es natural que nuestra decisión dependa de X.

Definicin 10 *Llamamos riesgo Bayesiano medio de una decisión* $d(X)$ *basada sobre un estadístico* X*, la esperanza de la función de perdida* $w(d(X),\theta)$ *donde* θ *tiene la distribución* a priori Q*:*

$$\begin{aligned}
W_Q(d(X)) &= E\left[w(d(X),\theta)\right] \\
&= \int w(d(x),\theta)\ p(x,\theta)\ dx\ d\theta \\
&= \int w(d(x),\theta)\ p(x|\theta)\ q(\theta)\ dx\ d\theta
\end{aligned}$$

y denotamos $d_Q(X)$, *la estrategia bayesiana con información muestral (o* a posteriori), *la decisión que minimiza* W_Q.

Teorema 11 *Si la función de perdidas es cuadrática* $w(d,\theta) = (d-\theta)^2$ *la estrategia Bayesiana* a posteriori *es la media* a posteriori *de* $\theta : d_Q(X) = E(\theta|X)$. *Además, tenemos que* $W_Q(d_Q(X)) = E_X\left[var(\theta|X)\right]$.

Tabla de los riesgos para una función de perdidas cudrática:

Conocimiento de θ	Mejor decisión	Riesgo mínimo
Perfecto	$d_\theta = \theta$	$w(d,\theta) = 0$
a priori: Q	$d_Q = E_Q(\theta)$	$\overline{w}_Q(d_Q) = \text{var}_Q(\theta)$.
a priori $+X$	$d_Q(X) = E(\theta\|X)$	$W_Q(d_Q(X)) = E_X\left[\text{var}(\theta\|X)\right]$

Ejemplo 18 *Una máquina esta ajustada para producir varillas de longitud θ, su producción es, debida a su imperfección, una secuencia de varillas de longitud $X_1, X_2, ..., X_n$, con distribución $N(\theta, \sigma^2)$. La varianza σ^2 es conocida (característica de la máquina).*
El estadístico sabe como fue ajustada la máquina : querrían ajustarla sobre el valor θ pero solo pudieron hacerlo solo con una cierta imprecisión. Eso lleva a admitir que θ tiene una distribución $N(m, \tau^2)$: conocemos la distribución a priori de θ.
El papel del estadístico es aquí predecir el valor de θ a partir de la observación de la producción de varillas.. Su error de predicción se mide por la función de perdidas $w(d, \theta) = (d - \theta)^2$.

- Si observamos directamente θ, la mejor decisión será siempre $d = \theta$.

- Pero aquí no observamos θ. Si no disponemos de la muestra de datos, nuestra decisión debe minimizar el riego bayesiano:

$$\begin{aligned} \overline{w}_Q(d) &= E_Q (d - \theta)^2 \\ &= \text{var}_Q(\theta) + (d - m)^2 \end{aligned}$$

  Por tanto, sin observación previa, la mejor decisión es $d_Q = m$: media a priori de θ. Para esta decisión, tenemos que $\overline{w}_Q(d_Q)$ =$\text{var}_Q(\theta) = \tau^2$.

- Ahora, suponemos que el estadístico dispone de las dos fuentes de información sobre θ : la distribución a priori $Q = N(m, \tau^2)$ y la muestra $X = (X_1, X_2, ..., X_n)$ que proviene condicionalmente a θ de una normal $N(\theta, \sigma^2)$. Con estas dos fuentes de información, la mejor predicción de θ es su esperanza condicionada por X. Tenemos que

$$f(\theta, x_1, x_2, ..., x_n) = C \exp \left(-\frac{1}{2\tau^2}(\theta - m)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \theta)^2 \right)$$

  y se puede deducir que

$$P(\theta | X_1, X_2, ..., X_n) = N \left(\frac{\tau^2 \sum X_i + \sigma^2 m}{n\tau^2 + \sigma^2}, \frac{\tau^2 \sigma^2}{n\tau^2 + \sigma^2} \right).$$

  Por tanto, la mejor predicción de θ será

$$d_Q(X) = \frac{\tau^2 \sum X_i + \sigma^2 m}{n\tau^2 + \sigma^2}$$

Para esta decisión el riesgo bayesiano medio será

$$\begin{aligned} W_Q(d_Q(X)) &= E_X\left[var(\theta|X)\right] \\ &= E_X\left[\frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}\right] = \frac{\tau^2\sigma^2}{n\tau^2+\sigma^2} \end{aligned}$$

que decrece hacia 0 cuando el tamaño muestral n crece hacia el infinito.