Introducción a la estadística bayesiana:
Un enfoque aplicado

Practicamente en cualquier área del conocimiento el investigador se enfrenta con la necesidad de analizar datos que presentan vari-abilidad. Lo anterior, con diversos objetivos. Los cuales van desde la simple descripción del fenómeno real que generó los resulta-dos, hasta la propuesta de modelos matemáticos que expliquen las relaciones de dependencia existentes entre las variables invo-lucradas. La Estadística permite analizar y modelar variables que presentan incertidumbre. En particular, el enfoque Bayesiano de la Estadística es una metodología basada en un conjunto de axiomas que permite realizar inferencias coherentes para analizar variables que presentan variabilidad. En este taller, se presentan los elementos esenciales de la metodología Bayesiana de la estadística, se describe cómo plantear y resolver problemas estadísticos empleando teoría de decisiones y se muestra cómo se puede aplicar esta metodología al análisis de datos que pre-sentan variabilidad.

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

6 Coloquio del Departamento de Matemáticas

Introducción a la estadística bayesiana: Un enfoque aplicado

Gabriel Nuñez-Antonio
Alberto Castillo Morales

Metepec, Atlixco, Puebla
Enero de 2014

Introducción a la Estadística Bayesiana:
Un enfoque aplicado

Gabriel Nuñez-Antonio
Departamento de Matemáticas, UAM-I
Área de Probabilidad y Estadística
gab.nunezantonio@gmail.com

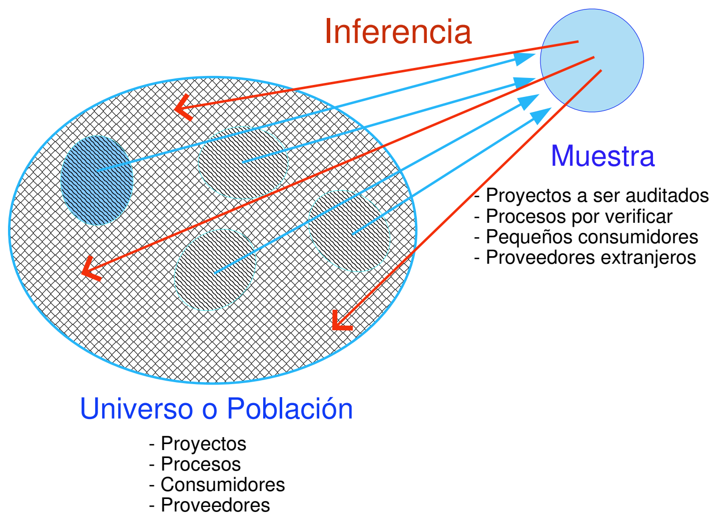
Alberto Castillo Morales
Departamento de Matemáticas, UAM-I
Área de Probabilidad y Estadística
ac@xanum.uam.mx

Octubre, 2013

Contenido

1 Introducción

- El **Objetivo** de la estadística, y en particular de la estadística Bayesiana, es proporcionar una metodología para analizar adecuadamente la información con la que se cuenta (*análisis de datos*) y decidir de manera razonable sobre la mejor forma de actuar (*toma de decisiones*)



- La **toma de decisiones** es un aspecto primordial en la vida de cualquier persona, por ejemplo:
  - Un médico debe de tomar decisiones constantemente en un ambiente de incertidumbre.
    * Decisiones sobre el diagnóstico más verosímil.
    * La oportunidad de una intervención quirúrgica.
    * El tratamiento más adecuado o la eficacia de un programa de inmunización.

- La **metodología estadística clásica** se puede ver como un conjunto de recetas que resultan apropiadas en determinados casos y bajo ciertas condiciones.

- Sin embargo, existe una **metodología unificada y general** que se deriva de analizar el proceso lógico que debe de seguirse para tomar una decisión (Teoría de decisión), y que incluye como caso particular al conjunto de procedimientos clásicos.

- La estadística está basada en la **teoría de probabilidades**. Formalmente la probabilidad es una función de valor real que cumple con ciertas condiciones. En términos prácticos puede entenderse como una medida o cuantificación de la incertidumbre.

- Aunque la definición de una medida de probabilidad es única, han existido (a lo largo de los siglos) diversas interpretaciones o enfoques: *clásica, frecuentista, subjetiva, etc.*

- **Enfoque Clásico**. Este enfoque considera que el fenómenos aleatorio produce resultados (eventos) igualmente verosímiles (posibles) y establece como medida de probabilidad el cociente entre los *casos favorables* y los *casos totales*:

$$P(A) = N(A)/N(\Omega)$$

- **Enfoque Frecuentista**. Este enfoque supone que el fenómeno o experimento aleatorio se puede repetir un número infinito de veces bajo condiciones similares y propone como medida de probabilidad el límite de la proporción de veces que ocurrió el evento de interés. Es decir:

$$P(A) = \lim_{N \to \infty} \frac{N(A)}{N(\Omega)}$$

- **Enfoque Subjetivo**. En este enfoque, la probabilidad es simplemente una medida *asignada por el decisor* de la incertidumbre asociada a un evento. Es decir, es sólo un juicio personal sobre la verosimilitud de la ocurrencia de cierto suceso o evento.

- **La metodología Bayesiana** está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el **Teorema de Bayes**

Reverendo Tomas Bayes

(Londres, 1702-Tunbridge Wells, 1761) Matemático británico. Estudió el problema de la determinación de la probabilidad de las causas a través de los efectos observados.

2 La controversia frecuentista-Bayesiana

- Mientras la probabilidad ha sido tema de estudio por cientos de años (muchos matemáticos son recordados por los consejos a los grandes ricos de cómo maximizar las ganancias en juegos de azar).
- La estadística es relativamente un campo joven:
  - La regresión lineal aparece en el trabajo de Francis Galton a finales del siglo XIX.
  - Con Karl Pearson se completa el siglo al incorporarse medidas de bondad de ajuste.
  - El campo realmente floreció hasta los 1920s y 1930s, cuando R.A. Fisher desarrolló la noción de verosimilitud para el problema de estimación y, Jerzy Neyman y Egon Pearson, las bases para las pruebas de hipótesis clásicas.
- La segunda Guerra Mundial propició un desarrollo repentino y acelerado en la investigación estadística.
  - Lo cual generó una amplia variedad de complejos problemas aplicados y la creación de las primeras fundaciones gubernamentales relevantes para su solución en los Estados Unidos y Gran Bretaña.
- En contraste, los métodos Bayesianos son más viejos y se remontan al artículo original de 1763 del Rev. Thomas Bayes, un ministro y matemático amateur.
  - El área generó algún interés para Laplace, Gauss y otros en el siglo XIX, pero la aproximación Bayesiana fue ignorada (o activamente opuesta) por los estadísticos de principios del siglo XX.
  - Afortunadamente, durante este periodo, varios prominentes no-estadísticos, entre los más notables Harold Jeffreys (un físico) y Arthur Bowley (un econometrista), continuaron la divulgación y promovieron el interés y los beneficios de las ideas Bayesianas.

- Así, a principios de los 1950s estadísticos como L.J. Savage, Bruno de Finetti, Dennis Lindley, Jack Kiefer, y muchos otros empezaron a abogar por los métodos Bayesianos como remedios para ciertas deficiencias en la teoría clásica.

- A continuación se presentan un par de ejemplos que ilustran tales deficiencias en estimación por intervalos y pruebas de hipótesis.

Ejemplo 2.1 *Ejemplo. Suponga $X_i \sim N(\theta, \sigma^2)$, $i = 1, ..., n$ i.i.d y que el objetivo es obtener un estimador por intervalo del 95% para la media poblacional θ. Cuando n es suficientemente grande (algunos dicen, mayor a 30) una aproximación clásica debería utilizar el intervalo de confianza*

$$\delta(\boldsymbol{x}) = \bar{x} \pm 1.96 s/\sqrt{n}$$

- *Este intervalo tiene la propiedad de que, en promedio sobre aplicaciones sucesivas, $\delta(\boldsymbol{x})$ fallará en capturar la verdadera media θ sólo 5% de las veces.*

- *Una interpretación alternativa es, antes de que cualquier dato sea obtenido, la probabilidad que el intervalo contenga el verdadero valor es 0.95. Esta propiedad es atractiva en el sentido de que se mantiene para todos los valores verdaderos de θ y σ^2*

- *Por otro lado, su uso en cualquier contexto de análisis de datos es algo difícil de explicar y entender:*

  - *Después de registrar los datos y calcular $\delta(\boldsymbol{x})$, el intervalo contiene o no al verdadero valor de θ; su probabilidad no es 0.95. Es* 0 *o* 1*.* ***¡Oh no!***

  - *Después de observar $\boldsymbol{x}$, una oración como "El verdadero valor de θ tiene una posibilidad de caer en $\delta(\boldsymbol{x})$" no es válido, aunque mucha gente (incluyendo estadísticos independientemente de su enfoque filosófico) interpreta un intervalo de confianza en esta forma.*

  - *Así, para los frecuentistas, "95%" no es una probabilidad condicional, es una etiqueta asociada con el intervalo para indicar, ya sea, que tan probable es antes de que lo evaluemos, o como se comportaría en un muestreo repetitivo.*

  - *Un intervalo frecuentista al 99% debería ser más amplio, un intervalo al 90% más angosto, pero condicionalmente sobre $\delta(\boldsymbol{x})$, todos tendrán una probabilidad de 0 o 1, de contener al verdadero valor.* **! !**

- *En contraste, los intervalos de confianza Bayesianos (conocidos como "conjuntos creíbles") son libres de estas interpretaciones frecuentistas no claras.*

  - *Por ejemplo, condicional sobre los datos observados, la probabilidad de que θ esté en el intervalo de credibilidad del 95% es de 0.95. Por supuesto esta interpretación natural es producto de pagar el precio de una distribución inicial (o a priori) para θ, así uno debe establecer estructuras adicionales para obtener esta naturalidad de las interpretaciones.*

  La estructura de pruebas de hipótesis de Neyman-Pearson también puede dirigir a resultados peculiares.

Ejemplo 2.2 *Considere el siguiente experimento simple, originalmente sugerido por Lindley y Phillips (1976), y reimpreso varias veces.*

- *Se tienen 12 lanzamientos independientes de una moneda y se observan 9 soles y 3 águilas. Se desea probar la hipótesis nula $H_0 : \theta = \frac{1}{2}$ contra la hipótesis alternativa $H_a : \theta > \frac{1}{2}$, donde θ es la verdadera probabilidad del sol. Dado sólo esta información, existen dos posibles modelos muestrales como candidatos.*

1. *Binomial: El número de $n = 12$ lanzamientos se fija con anterioridad y la cantidad aleatoria X es el número de soles observado en los n lanzamientos.*

   Entonces, $X \sim Bin(12, \theta)$ y la verosimilitud está dada por:

$$L_1(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3.$$

2. *Binomial Negativa: La recolección de los datos involucra lanzamientos de la moneda hasta que la tercer águila aparece. Aquí, la cantidad aleatoria X es el número de soles requeridos para completar el experimento, así que $X \sim BinNeg(r = 3, \theta)$, con función de verosimilitud dada por*

$$L_2(\theta) = \binom{r+x-1}{x} \theta^x (1-\theta)^r = \binom{11}{9} \theta^9 (1-\theta)^3.$$

Bajo cualquiera de esta dos alternativas , uno puede calcular los valores-p correspondientes a la región de rechazo, "Rechazar H_0 si $X \geq c$".

- *Para la distribución binomial se tendría:*

$$\alpha_1 = P_{\theta=\frac{1}{2}}(X \geq 9) = \sum_{j=9}^{12} \binom{12}{j} \theta^j (1-\theta)^{12-j} = 0.075$$

- *Mientras que para la binomial negativa:*

$$\alpha_2 = P_{\theta=\frac{1}{2}}(X \geq 9) = \sum_{j=9}^{\infty} \binom{2+j}{j} \theta^j (1-\theta)^3 = 0.0325$$

- *Así, usando el "usual" nivel de error tipo I de $\alpha = 0.05$, podemos ver que los dos modelos supuestos conducen a diferentes decisiones.*

- *Sin embargo, no existe información en el contexto del problema que ayude en la determinación del modelo, así que no es claro cual análisis frecuentista debería considerarse como "correcto".*

- *En cualquier caso, suponiendo que confiamos en el modelo estadístico, es difícil identificar (razonablemente) cómo la manera en que se monitoreo el experimento pueda dar soporte a nuestra decisión; seguramente sólo sus resultados son relevantes* **!**

- *En realidad las funciones de verosimilitud ofrecen una historia inconsistente, desde que L_1 y L_2 sólo difieren por una constante multiplicativa que no depende de θ.*

Una explicación Bayesiana de lo que esta erróneo en el ejemplo anterior sería que el enfoque de Neyman-Pearson permite que resultados "no-observados" afecten la decisión de rechazar H_0.

- Estos es, la probabilidad de valores de X más extremos que 9 son usados como evidencia contra H_0 en cada caso, **aunque estos valores no ocurrieron**.

- Lo anterior, es una violación de un axioma estadístico conocido como *Principio de Verosimilitud*: Una vez que los valores de los datos $\boldsymbol{x}$ han sido observados, la función de verosimilitud $L(\theta|\boldsymbol{x})$ contiene toda la información experimental relevante ofrecida por $\boldsymbol{x}$ sobre el parámetro desconocido θ.

- En los ejemplos anteriores, L_1 y L_2 son proporcionales una de la otra como funciones de θ, así, son equivalentes en términos de información experimental.

Algunos estadísticos intentan defender los resultados del ejemplo 2 argumentando que todos los aspectos del diseño de un experimento son piezas relevantes de información aún después de que los datos han sido recolectados, o quizá que el Principio de Verosimilitud es en sí defectuoso (o al menos no debería ser considerado sacrosantamente)

Si los métodos Bayesianos ofrecieron una solución a estos y otros inconvenientes del enfoque frecuentista estas fueron publicadas hace varios años.

Parece sorprendente que la metodología Bayesiana no tomara mayor ventaja en la práctica estadística, sólo recientemente.

¡ ¿Qué pasó? !

1°. Los defensores duros iniciales de los métodos Bayesianos defendían fuertemente la *"subjetividad"* del conocimiento inicial, en contraposición al enfoque clásico.

2°. (Y quizá más importante desde un punto de vista aplicado)
La alternativa Bayesiana, mientras era simple teóricamente, requería evaluaciones de integrales complejas aún en cierto problemas básicos.

    - La falta de rapidez, los altos costos de los recursos computacionales y los problemas teóricos asociados limitaron el crecimiento del análisis de datos reales desde un punto de vista Bayesiano.

Finalmente, este *boom* se dio en los 1980s, gracias a un grupo más objetivo de Bayesianos con acceso a recursos computacionales baratos y veloces.

3 Estructura de un problema de Decisión

- ¿Qué es un problema de decisión?. Nos enfrentamos a un problema de decisión cuando debemos elegir entre dos o más formas de actuar.

- La mayoría de las decisiones cotidianas son relativamente fáciles de tomar. Por ejemplo, cuando elegimos una película en cartelera o el platillo fuerte en algún restaurante.

- Sin embargo, existen problemas de decisión en donde las **consecuencias** son importantes y exigen una sería reflexión. Por ejemplo: la decisión de casarse con cierta persona o la de cambiar de trabajo.

- **La Teoría de Decisiones** propone una forma (metodología) de tomar decisiones basada en unos cuantos principios básicos sobre la **elección coherente** entre diversas alternativas (**opciones**).

3.1 Elementos de un problema de decisión

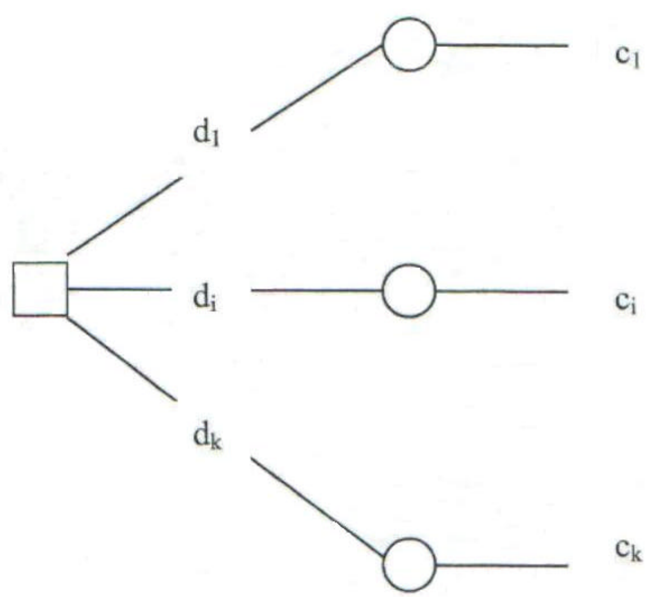
1. **Espacio de Decisiones ($\boldsymbol{D}$).** Es el conjunto de posibles alternativas, el cual debe ser **exhaustivo** (agotar todas las posibilidades que en principio parezcan razonables) y **mutuamente excluyente**.
$$\boldsymbol{D} = \{d_1, ..., d_k\}$$

2. **Espacio de Consecuencias ($\boldsymbol{C}$).** Es el conjunto de consecuencias posibles y describe las consecuencias de elegir cierta decisión.
$$\boldsymbol{C} = \{c_1, ..., c_m\}$$

3. **Relación de Preferencia ($\lesssim$)** . Relación de preferencia entre las distintas consecuencias. Se define de manera que
$$c_1 \lesssim c_2 \text{ si } c_2 \text{ es preferido a } c_1.$$

- Un problema de decisión se puede representar mediante un **árbol** .

Si la decisión se realiza **sin incertidumbre** a cada decisión le corresponde una consecuencia segura.



Ejemplo 3.1 *Un estudiante de preparatoria se enfrenta al problema de decidir qué carrera estudiar.*

$\boldsymbol{D} = \{d_1, d_2, d_3, d_4\}$, *donde* $d_1 \equiv$ *arquitectura,* $d_2 \equiv$ *ingeniería industrial,* $d_3 \equiv$ *actuaría y* $d_4 \equiv$ *matemáticas.*

$\boldsymbol{C} = \{c_1, c_2, c_3, c_4\}$, *donde* $c_1 = 20,000$, $c_2 = 25,000$, $c_3 = 30,000$, $c_4 = 15,000$.

$\lesssim = \{3^0, 2^0, 1^0, 4^0\}$ *orden de preferencia.*

- **El objetivo** de un problema de decisión es determinar **la mejor** decisión de entre un conjunto de alternativas.
- Para lo anterior, el decisor debe ser capaz de **comparar** y establecer un orden ($\lesssim$) entre las consecuencias.
- Es poco común que se conozcan con certeza todas las posibles consecuencias de tomar una decisión, por lo que un problema de decisión en general, se plantea bajo un **ambiente de incertidumbre** .

$$\Longrightarrow$$

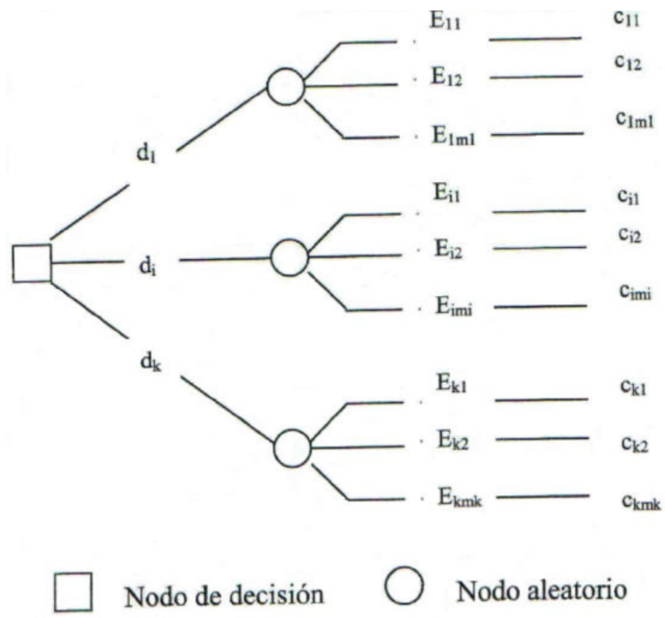
- Elemento extra de un problema de decisión:

  4. **Espacio de Eventos inciertos ($\boldsymbol{E}$).** Es el conjunto de los eventos o sucesos inciertos relevantes al problema de decisión.

$$\boldsymbol{E} = \{E_1, ..., E_m\}$$

- Árbol de Decisión en **ambiente de incertidumbre** .

Si la decisión se realiza en ambiente de incertidumbre no se tiene información completa sobre las consecuencias de tomar cierta decisión.



Ejemplo 3.2 *Un médico debe decidir si realizar una operación a una persona que se cree puede tener un tumor, o recurrir a una determinada medicación. Si el paciente no tiene un tumor su esperanza de vida se estima en 20 años. Si lo tiene, se opera y si sobrevive a la operación, le dan 10 años de vida, y si se tiene el tumor y no se opera, sólo le dan 2 años de vida.*

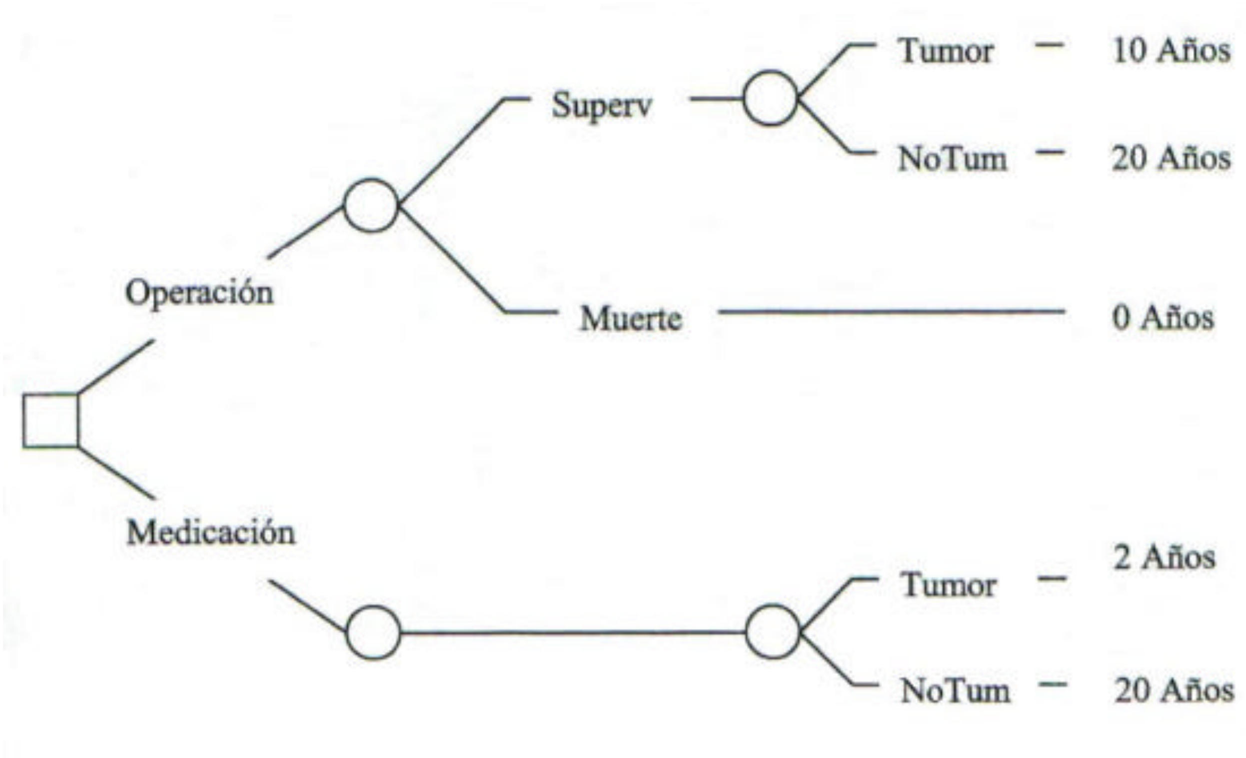
$\boldsymbol{D} = \{d_1, d_2\}$, *donde* $\quad d_1 =$ *operar,* $d_2 =$ *medicar.*

$\boldsymbol{E} = \{E_{11}, E_{12}, E_{13}, E_{21}, E_{23}\}$, *donde*

$$E_{11} = \textit{supervivencia / tumor},\ E_{12} = \textit{supervivencia / no tumor},$$
$$E_{13} = \textit{muerte},\ E_{21} = \textit{tumor},\ E_{22} = \textit{no tumor}.$$

$\boldsymbol{C} = \{c_{11}, c_{12}, c_{13}, c_{21}, c_{23}\}$, *donde*

$$c_{11} = 10,\ c_{12} = 20,\ c_{13} = 0,\ c_{21} = 2,\ c_{22} = 20.$$



- Observaciones:
  - En la práctica, la mayoría de los problemas de decisión tienen una estructura más compleja. Por ejemplo, decidir si realizar o no una inversión financiera en instrumentos de renta variable, y en caso afirmativo decidir si invertir en la bolsa mexicana de valores, en renta variable nacional o en renta variable extranjera (***problemas secuenciales de decisión***).
  - Frecuentemente, el **conjunto de eventos inciertos es el mismo** para cualquier decisión que se tome. En este caso, el problema de decisión se puede representar mediante una tabla.

	$E_1 \ldots E_j \ldots E_m$
d_1	$C_{11} \ldots C_{1j} \ldots C_{1m}$
$\vdots$	$\vdots \quad \vdots \quad \vdots$
d_i	$C_{i1} \ldots C_{ij} \ldots C_{im}$
$\vdots$	$\vdots \quad \vdots \quad \vdots$
d_k	$C_{k1} \ldots C_{kj} \ldots C_{km}$

Ejemplo 3.3 *Se dispone de un automóvil y de una motocicleta para ir a cita en el centro de la ciudad, aunque siempre existe la opción de ir a pie. ¿Qué alternativa elegir?. La moto es*

más rápida, sobre todo si hay tráfico y puede estacionarse con facilidad, pero resulta incómoda si hace mal tiempo. El coche aunque es más cómodo consume más gasolina y hay que dejarlo en un estacionamiento a cierta distancia del lugar de destino. Supongamos que no hay transporte público disponible y el costo de un taxi es excesivo.

$\boldsymbol{D} = \{d_1, d_2, d_3\}$, *donde* $d_1 =$ *moto*, $d_2 =$ *coche y* $d_3 =$ *pie.*

$\boldsymbol{E} = \{E_1, E_2\}$, *donde* $E_1 =$ *llueve y* $E_2 =$ *no llueve.*

$\boldsymbol{C} = \{c_{11}, c_{12}, c_{21}, c_{22}, c_{31}, c_{32}\}$, *donde las* c_{ij} *se describen en la siguiente tabla*

	$E_1 \equiv$ *llueve*	$E_2 \equiv$ *no llueve*
$d_1 \equiv$ *moto*	$c_{11} =$ *Poco tiempo, trayecto incómodo, costo bajo.*	$c_{12} =$*Poco tiempo, trayecto agradable, costo bajo.*
$d_2 \equiv$*coche*	$c_{21} =$ *Algo de tiempo, manejo con lluvia, costo moderado.*	$c_{22} =$*Algo de tiempo, manejo sin lluvia, costo moderado.*
$d_3 \equiv$ *pie*	$c_{31} =$ *Mucho tiempo, caminata con lluvia, sin costo.*	$c_{32} =$*Mucho tiempo, caminata sin lluvia, sin costo.*

El mismo problema también se puede representar mediante un árbol de decisión.

- El **Objetivo** de un problema de decisión en ambiente de incertidumbre consiste entonces en elegir *la mejor* decisión, d_i, de $\boldsymbol{D}$ sin saber cuál de los eventos inciertos E_{ij} de $\boldsymbol{E_i}$ ocurrirá.
- Aunque los sucesos que componen cada $\boldsymbol{E_i}$ son inciertos, en el sentido de que no sabemos cuál de ellos tendrá lugar, en general se tiene una idea sobre la probabilidad de cada uno de ellos. Por ejemplo, si un individuo hoy cumplió 40 años, ¿cuál de los siguientes eventos es más probable?
  - Que viva 10 años más
  - Que muera en un mes.
  - Que llegue a los 100 años

3.2 Cuantificación de las consecuencias y de los eventos inciertos

- Alguna veces resulta difícil ordenar nuestras preferencias sobre las distintas consecuencias posibles. Tal vez resulta más fácil asignar una utilidad (o pérdida) a cada una de las consecuencias y ordenarlas de acuerdo a sus utilidades.

Consecuencias – – –– → Utilidades

– Ganar mucho dinero y tener poco tiempo disponible ⇒ \$

– Ganar regular de dinero y tener regular tiempo disponible ⇒ \$\$

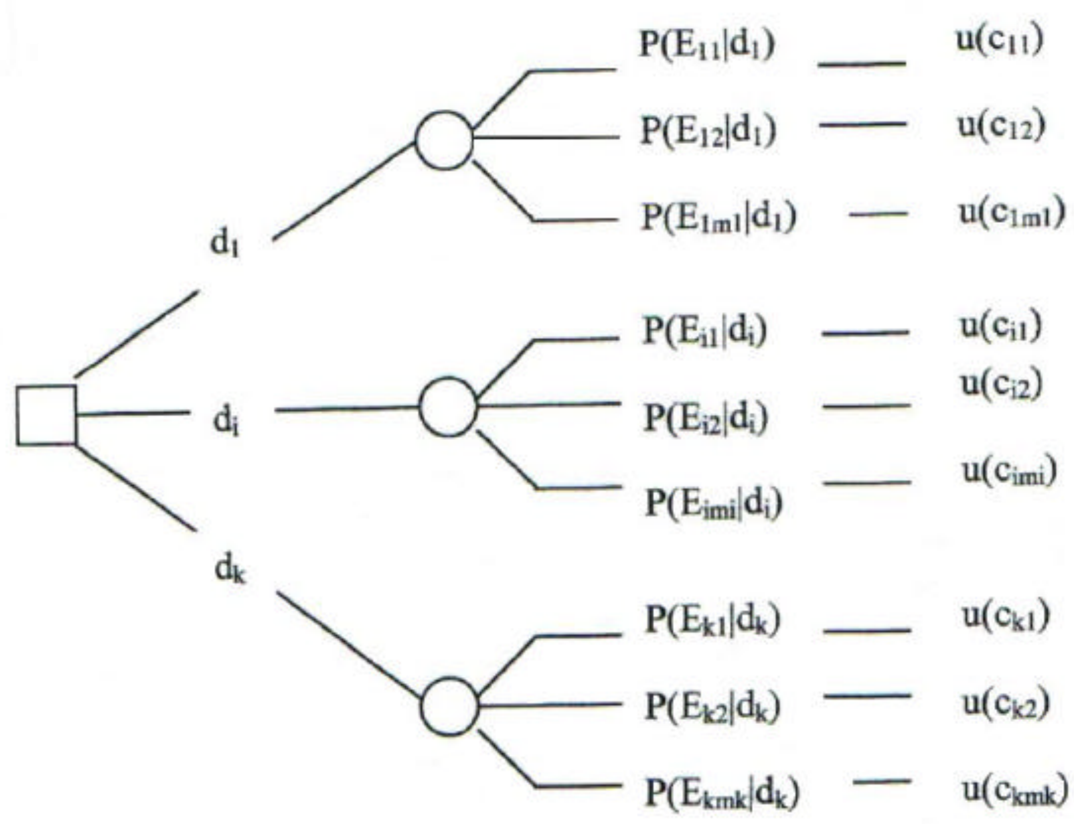
– Ganar poco dinero y tener mucho tiempo disponible ⇒ \$\$\$

- En general, es posible cuantificar las preferencias del decisor entre las distintas consecuencias a través de una **función de utilidad** de manera que

$$c_{ij} \lesssim c_{i_* j_*} \quad \Leftrightarrow \quad \boldsymbol{u}(c_{ij}) \leq \boldsymbol{u}(c_{i_* j_*})$$

- De la misma manera, la información que el decisor tiene sobre la posible ocurrencia de los eventos inciertos puede ser cuantificada a través de una **función de probabilidad** sobre el espacio de eventos $\boldsymbol{E}$: $P(E_{ij})$.

- Por lo anterior, es posible representar el árbol de decisiones de la siguiente manera:



3.3 Criterios para resolver un problema de Decisión

¿Cómo tomar la *mejor* decisión?

Si de alguna manera fuéramos capaces de deshacernos de la incertidumbre podríamos ordenar nuestras preferencias de acuerdo con las utilidades de cada decisión. *La mejor decisión* sería la que tenga la máxima utilidad.

3.3.1 Estrategias o Criterios

1. *Optimista.* Elegir la mejor consecuencia de cada opción y elegir ...
2. *Pesimista o minimax.* Elegir la peor consecuencia de cada opción y elegir ...
3. *Consecuencias más probable o condicional.* Elegir la consecuencia más probable de cada opción y...
4. *Utilidad Esperada.* Elegir una consecuencia esperada para cada opción y ...

⋆ Cualquiera que sea la estrategia tomada, la *mejor* decisión es aquella que maximice la utilidad del árbol "sin incertidumbre"

Ejemplo 3.4 *En unas elecciones parlamentarias en la Gran Bretaña competían los partidos Conservador y Laboral, y una casa de apuestas ofrecía las siguientes posibilidades:*

a) A quien apostara a favor del partido Conservador, la casa estaba dispuesta a pagar 7 libras por cada 4 que apostase si el resultado favorecía a los conservadores, en caso contrario el apostador perdía su apuesta.

b) A quien apostara a favor del partido Laboral, la casa estaba dispuesta a pagar 5 libras por cada 4 que apostase si ganaban los laboristas, en caso contrario el apostador perdía su apuesta.

- *¿A qué partido apostar?*

  $\boldsymbol{D} = \{d_1, d_2\}$, *donde*

$$\begin{aligned} d_1 &= \textit{Apostar al partido conservador} \\ d_2 &= \textit{Apostar al partido laboral} \end{aligned}$$

  $\boldsymbol{E} = \{E_1, E_2\}$, *donde*

$$\begin{aligned} E_1 &= \textit{Que gane el partido conservador} \\ E_2 &= \textit{Que gane el partido laboral} \end{aligned}$$

  Sea

$$\begin{aligned} \pi &= P(\textit{gane el partido conservador}) = P(E_1) \\ 1-\pi &= P(\textit{Que gane el partido laboral}) = P(E_2) \end{aligned}$$

$\boldsymbol{C} = \{c_{11}, c_{12}, c_{21}, c_{22}\}$.

Si la apuesta es de k libras, entonces

$$\begin{aligned} c_{11} &= -k + (7/4)k = \tfrac{3}{4}k \\ c_{12} &= -k \\ c_{21} &= -k \\ c_{22} &= -k + (5/4)k = \tfrac{1}{4}k \end{aligned}$$

Si las consecuencias se valoran en términos de dinero, entonces, la utilidad es proporcional al dinero, i.e., $u(c_{ij}) = c_{ij}$.

- **Caso 1 :** $\pi = 1/2$

$P(E)$	$1/2$	$1/2$
$u(d, E)$	E_1	E_2
d_1	$(3/4)k$	$-k$
d_2	$-k$	$(1/4)k$

1. *Optimista: d_1. Apostar al partido conservador.*
2. *Pesimista: d_1 ó d_2, da igual cualquiera de las dos.*
3. *Consecuencia más probable: d_1 ó d_2, da igual cualquiera de las dos.*
   - *Si se toma a E_1 como "seguro" $\Rightarrow d_1$*
   - *Si se toma a E_2 como "seguro" $\Rightarrow d_2$.*
4. *Utilidad esperada: d_1. Apostar al partido conservador.*
   - $E(u(d_1)) = (1/2)(3/4)k + (1/2)(-k) = (-1/8)k$
   - $E(u(d_2)) = (1/2)(-k) + (1/2)(1/4)k = (-3/8)k$.

- **Caso 2 :** $\pi = 1/4$ *(Ejercicio 1 (tarea)...)*

Observación: *Existen ciertos valores de π para los cuales se opta por d_1 y otros para los cuales se opta por d_2 como la mejor opción.*

- **Caso General :** $\pi \in [0, 1]$.

$P(E)$	π	$1 - \pi$
$u(d, E)$	E_1	E_2
d_1	$(3/4)k$	$-k$
d_2	$-k$	$(1/4)k$

1. Optimista: d_1. Apostar al partido conservador.

2. Pesimista: d_1 ó d_2, da igual cualquiera de las dos.

3. Consecuencia más probable: d_1 ó d_2, (dependiendo)

  – *Si $\pi > 1/2$ se toma a E_1 como "seguro" $\Rightarrow d_1$.*

  – *Si $\pi \leq 1/2$ se toma a E_2 como "seguro" $\Rightarrow d_2$.*

4. Utilidad esperada: d_1 ó d_2, (dependiendo).

Las utilidades esperadas son:

$$E(u(d_1)) = \pi(3/4)k + (1-\pi)(-k) = \{(7/4)\pi - 1\}k$$

$$E(u(d_2)) = \pi(-k) + (1-\pi)(1/4)k = \{(1/4) - (5/4)\pi\}k.$$

Entonces, la mejor decisión sería:

- $E\{u(d_1)\} > E\{u(d_2)\} \quad \Leftrightarrow \quad \pi > \frac{5}{12} \quad \Rightarrow \quad d_1$
- $E\{u(d_1)\} < E\{u(d_2)\} \quad \Leftrightarrow \quad \pi < \frac{5}{12} \quad \Rightarrow \quad d_2$
- $E\{u(d_1)\} = E\{u(d_2)\} \quad \Leftrightarrow \quad \pi = \frac{5}{12} \quad \Rightarrow \quad d_1$ ó d_2.

Si definimos las funciones

$$g_1(\pi) = \{(7/4)\pi - 1\}k = E(u(d_1))$$

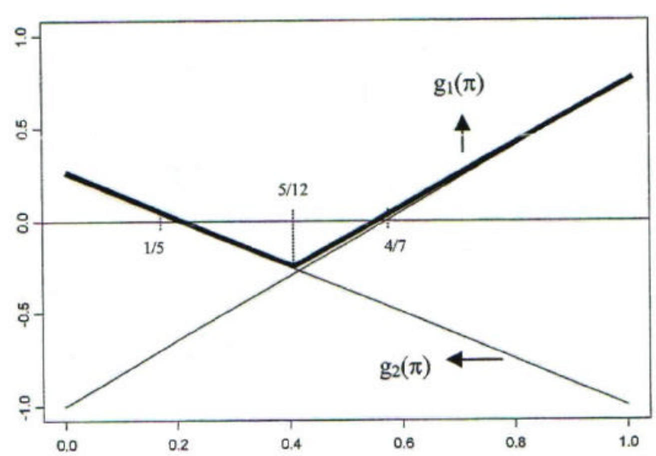
$$g_2(\pi) = \{(1/4) - (5/4)\pi\}k = E(u(d_2)).$$

Entonces, para $k = 1$, se tiene

- *La* **línea gruesa** *representa la mejor solución al problema de decisión dada por el criterio de la utilidad esperada.*

- *Observación: Si $\pi \in [\frac{1}{5}, \frac{4}{7}]$, la utilidad esperada de la mejor decisión es negativa* **!**

- *Pregunta: ¿Tú apostarías si $\pi \in [\frac{1}{5}, \frac{4}{7}]$?.*



Considera ahora el problema de decisión (general) con d_3, donde d_3 es no apostar. Es decir, $\boldsymbol{D} = \{d_1, d_2, d_3\}$.

En este caso las utilidades esperadas son:

$E(u(d_1)) = \pi(3/4)k + (1-\pi)(-k) = \{(7/4)\pi - 1\}k$

$E(u(d_2)) = (\pi(-k) + (1-\pi)(1/4)k = \{(1/4) - (5/4)\pi\}k.$

$E(u(d_3)) = \pi(0) + (1-\pi)(0) = 0.$

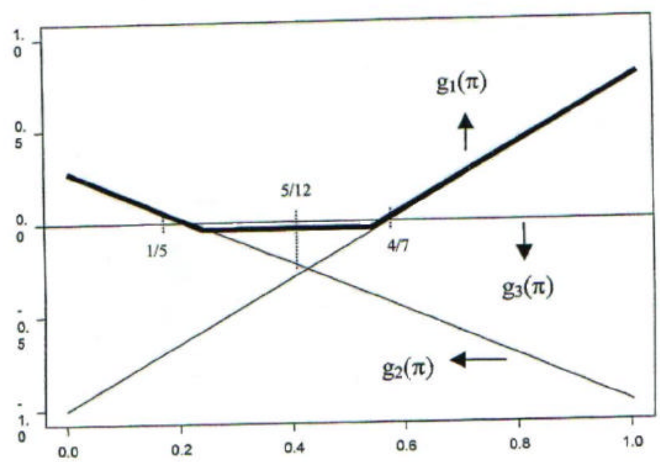
La mejor decisión sería la que maximice la utilidad esperada para distintos valores de $\pi \in [0,1]$. Si consideramos ahora las funciones

$$g_1(\pi) = \{(7/4)\pi - 1\}k = E(u(d_1))$$

$$g_2(\pi) = \{(1/4) - (5/4)\pi\}k = E(u(d_2))$$

$$g_3(\pi) = 0 = E(u(d_3)).$$

Entonces, para $k = 1$, se tiene



- *La* **línea gruesa** *representa la mejor solución al nuevo problema de decisión dada por el criterio de la utilidad esperada.*
- *Por lo tanto, la mejor decisión sería:*
  - *Si* $\pi < \frac{1}{5} \Rightarrow d_2$.
  - *Si* $\frac{1}{5} \leq \pi < \frac{4}{7} \Rightarrow d_3$.
  - *Si* $\pi > \frac{4}{7} \Rightarrow d_1$.

Observación: En este caso, la utilidad de la mejor decisión nunca es negativa.

Pregunta: ¿Cuál es el valor que la casa esperaría tuviera π? Intuitivamente, la casa esperaría que $\pi \in [\frac{1}{5}, \frac{4}{7}]$

Moraleja: Vale la pena incluir todas las opciones factibles en una problema de decisión, excepto las inadmisibles.

3.3.2 Inadmisibilidad

Se dice que una opción d_1 es **inadmisible** si existe otra opción d_2 tal que d_2 es al menos tan preferible como d_1 pase lo que pase (para cualquier suceso incierto) y existe un caso (suceso incierto) para el que d_2 es más preferible que d_1.

Ejemplo 3.5 *En una lotería que tiene mil números se sortea un premio mayor de \$5,000 pesos y se dan \$10 pesos (reintegro) a todos los boletos cuya última cifra coincide con la del primer premio. Supón que la utilidad del dinero es proporcional a su cantidad. Si el costo de cada billete es de \$10 pesos, ¿Comprarías tú un billete de lotería?, ¿Cuál debería ser el costo justo del billete?*

En este caso:

- $\boldsymbol{D} = \{d_1, d_2\}$, *donde*

$$\begin{aligned} d_1 &\equiv \text{Comprar un billete de lotería} \\ d_2 &\equiv \text{No comprar un billete de lotería} \end{aligned}$$

- $\boldsymbol{E} = \{E_1, E_2, E_3\}$, *donde*

$$\begin{aligned} E_1 &\equiv \text{Que me gane el premio mayor} \\ E_2 &\equiv \text{Que me gane un reintegro} \\ E_3 &\equiv \text{Que no me gane nada} \end{aligned}$$

La probabilidad de cada uno de estos sucesos inciertos es

$$\begin{aligned} P(E_1) &= \tfrac{1}{1000} = 0.001 \\ P(E_2) &= \tfrac{100-1}{1000} = 0.099 \\ P(E_3) &= 0.9(\textit{por completez}) \end{aligned}$$

- $\boldsymbol{C} = \{c_{11}, c_{12}, c_{13}, c_{21}\}$, *donde*

$$\begin{aligned} c_{11} &= 5000 - 10 = 4990 \\ c_{12} &= 10 - 10 = 0 \\ c_{13} &= 0 - 10 = -10 \\ c_{21} &= 0 \end{aligned}$$

Si la utilidad esperada es el dinero, i.e., $u(c_{ij}) = c_{ij}$, entonces:

1. *Optimista: d_1. Comprar.*
2. *Pesimista: d_2. No comprar.*

3. Consecuencia más probable: d_2, *(no comprar)*

4. Utilidad esperada: d_2, *(no comprar).*

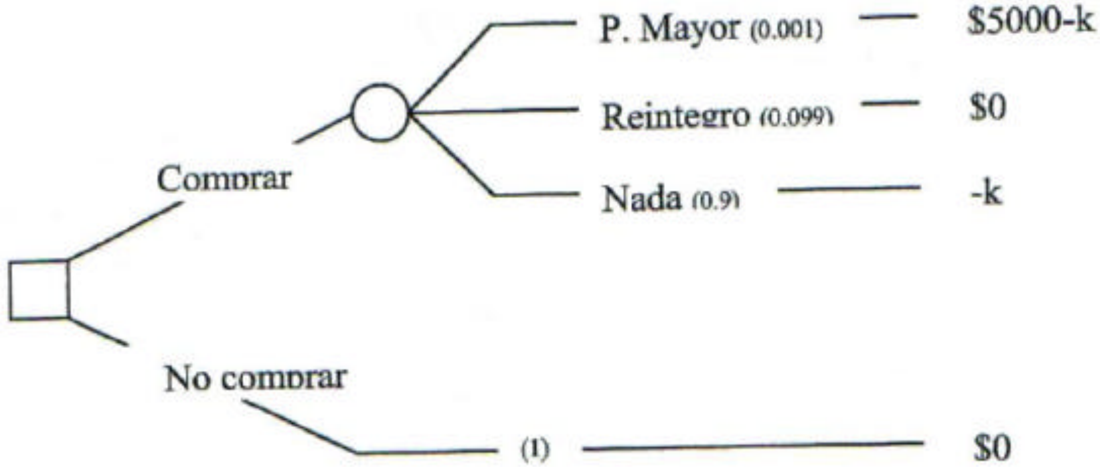
Las utilidades esperadas son:

- $E(u(d_1)) = (0.001)4990 + (0.099)0 + (0.9)(-10) = -4.01.$
- $E(u(d_2)) = (1)0.$

Observación: *La utilidad esperada de comprar un billete de lotería, desde el punto de vista del comprador es negativa, pero desde el punto de vista del vendedor, es positiva!*

¿Cómo determinar el *costo justo* ***del billete de lotería?.***

Si denotamos por k *el costo del billete, entonces tenemos el siguiente árbol de decisión*



Utilizando la estrategia de la utilidad esperada:

$$E(u(d_1)) = (0.001)(5000 - k) + (0.099)0 + (0.9)(-k) = 5 - 0.901k$$

$$E(u(d_2)) = (1)0 = 0$$

Finalmente,

$$E(u(d_1)) > E(u(d_2)) \Leftrightarrow 5 - (0.901)k > 0 \Leftrightarrow k < 5.55 \Rightarrow d_1$$

$$E(u(d_1)) < E(u(d_2)) \Leftrightarrow 5 - (0.901)k < 0 \Leftrightarrow k > 5.55 \Rightarrow d_2$$

$$E(u(d_1)) = E(u(d_2)) \Leftrightarrow k = 5.55 \Rightarrow d_1 \text{ ó } d_2.$$

Por lo tanto, el *precio justo* ***del billete de lotería es de \$5.55 pesos.***

Ejercicio 3.1 *El día de su cumpleaños número 20, un paciente es admitido en el hospital con síntomas que sugieren la enfermedad A (con probabilidad 0.4) o bien la enfermedad B (con probabilidad 0.6). Cualquiera que sea la enfermedad que en realidad padece, si no se trata morirá ese mismo día (con probabilidad 0.8) o bien sobrevivirá sin consecuencias para su salud (con probabilidad 0.2). El médico que recibe a este paciente tiene tres opciones (alternativas):*

1. *No administrar tratamiento alguno.*
2. *Tratar al paciente con el fármaco G.*
3. *Operar al paciente (Cirugía)*

Tanto la cirugía como la administración de la droga entrañan riesgos. Sin importar la enfermedad, el paciente puede morir durante la operación (con probabilidad 0.5). De la misma manera, sin importar la enfermedad, la droga puede ocasionar un problema renal mortal (con probabilidad 0.2).

Si el paciente sobrevive a los efectos adversos de la droga, y tenía la enfermedad A, entonces se puede curar (con probabilidad 0.9) o que no tenga efecto alguno. Si el paciente tenía la enfermedad B, seguro la droga no tiene efecto curativo.

Por otra parte, si el paciente sobrevive a la operación, esta lo cura con probabilidad 0.5 si padecía la enfermedad A (en caso contrario no tiene efecto). Si por el contrario, padecía la enfermedad B, la operación lo cura con probabilidad 0.6 y no tiene efecto con probabilidad 0.4.

En cualquier caso si se recupera, el paciente tiene una esperanza de vida de 50 años o más.
¿Qué tratamiento es más conveniente?, ¿Cuál es el valor de Bayes de la decisión óptima?

Ejercicio 3.2 *Una compañía de seguros debe decidir si crea una póliza de seguros que cubra completamente uno de los dos tipos de riesgos a los que estarán expuestas las nuevas compañías textiles que se instalarán en México el próximo año, o bien, crear una póliza que cubra los dos tipos de riesgos parcialmente. La compañía de seguros considera que lo más probable es que las nuevas compañías textiles estarán más expuestas al riesgo tipo 2.*

El espacio de decisiones en este problema es:

$d_1 =$ *crear una póliza que cubra completamente el riesgo tipo 1.*

$d_2 =$ *crear una póliza que cubra completamente el riesgo tipo 2.*

$d_3 =$ *crear una póliza que cubra los dos tipos de riesgo pero con una cobertura menor para ambos riesgos.*

El espacio de sucesos inciertos podría describirse como:

$$\begin{aligned}
\theta_1 &= \text{las nuevas compañías textiles estarán propensas al riesgo tipo 1.}\\
\theta_2 &= \text{las nuevas compañías textiles estarán propensas al riesgo tipo 2.}\\
\theta_3 &= \text{las nuevas compañías textiles estarán propensas tanto al riesgo}\\
&\quad\ \text{tipo 1 como al riesgo tipo 2.}
\end{aligned}$$

De acuerdo a la compañía de seguros se presenta la siguiente tabla de utilidades

	θ_1	θ_2	θ_3	*mín*
d_1	0.9	0.2	0.5	0.2
d_2	0.2	0.9	0.5	0.2
d_3	0.6	0.6	0.7	0.6
			máx	0.6

Analice este problema de decisión considerando todos los criterios revisados anteriormente.

Ejercicio 3.3 *Considerando un problema de decisión definido por la siguiente tabla. ¿Qué criterio o estrategia de decisión es mejor?*

	θ_1	θ_2	θ_3
$P(\theta = \theta_i)$	0.32	0.30	0.38
d_1	.1	.1	.4
d_2	.4	.4	.1

4 Tratamiento Axiomático de un Problema de Decisión

- Recordemos que un problema de decisión en ambiente de incertidumbre queda completamente especificado si se definen los elementos $\boldsymbol{D}, \boldsymbol{E}, \boldsymbol{C}$ y $\precsim$.
- Algunas definiciones quedan un poco ambiguas cuando se plantea un problema de decisión, son las siguientes:
  - Una vez que se establecieron las posibles consecuencias (para cada decisión), muchas veces no es claro para el decisor cuál de ellas prefiere.
  - Establecer o definir los eventos inciertos que pueden ocurrir para cada decisión, no es tarea fácil. Lo anterior, ya que siempre existe la duda si se contemplaron todas las posibilidades.
  - Aún cuando el decisor sea capaz de ordenar sus preferencias y se hayan contemplado todas los eventos inciertos posibles, ¿Qué criterio debemos emplear para resolver el problema de decisión?

4.1 Axiomas de Coherencia

- Los denominados *axiomas de coherencia* son una serie de principios que establecen las condiciones para que las ambigüedades anteriores se clarifiquen.
- **Axiomas de Coherencia** :
  - Comparabilidad.
  - Transitividad.
  - Sustitución y Dominancia.
  - Eventos de Referencia.
- Para entender mejor los axiomas de coherencia es necesario introducir una notación formal:
  - En general toda opción d_i se puede escribir en términos de todas sus posibles consecuencias dados los sucesos inciertos, es decir,

$$d_i = \{c_{ij}|E_{ij}, \quad j = 1, ..., m_i\}$$

1. **Comparabilidad** . Este axioma establece que al menos debemos ser capaces de expresar preferencias entre dos posibles decisiones y por lo tanto entre dos posibles consecuencias. Es decir, no todas las decisiones ni todas las consecuencias son iguales.
   - Para todo par de opciones d_1 y d_2 en $\boldsymbol{D}$, es cierta una y sólo una de las siguientes condiciones:
     - d_2 es más preferible que $d_1 \Leftrightarrow d_1 \lesssim d_2$.
     - d_1 es más preferible que $d_2 \Leftrightarrow d_2 \lesssim d_1$.
     - d_1 y d_2 son igualmente preferibles $\Leftrightarrow d_1 \sim d_2$

   Además es posible encontrar dos consecuencias c_* (la peor) y c^* (la mejor) tales que para cualquier otra consecuencia c, $c_* \leq c \leq c^*$

2. **Transitividad.** Este axioma establece que las preferencias deben de ser transitivas para no caer en contradicciones.
   - Si d_1, d_2 y d_3 son tres opciones cualesquiera y ocurre que $d_1 \lesssim d_2$ y $d_2 \lesssim d_3$, entonces, necesariamente $d_1 \lesssim d_3$. Análogamente, si $d_1 \sim d_2$ y $d_2 \sim d_3$, entonces $d_1 \sim d_3$.
3. **Sustitución y dominancia** .

- En virtud del axioma de sustitución, en una opción l se puede reemplazar una consecuencia por otra equivalente a ella. Es decir, si $c \sim l_1$, entonces $l_2 = \{c_1|A, c_2|\bar{A}\}$ es equivalente a $l_3 = \{l_1|A, c_2|\bar{A}\}$. Lo anterior, ya que si se da el evento A, l_2 da lugar a c_1 y l_3 da lugar a l_1 y por hipótesis $c_1 \sim l_1$, entonces si A $l_2 \sim l_3$. Por otro lado, si $\bar{A}$ l_2 y l_3 dan lugar a c_2 y en consecuencia $l_2 \sim l_3$. Por lo tanto, $l_1 \sim l_3$.
- Si d_1, d_2 son dos opciones cualesquiera y E es un evento incierto y sucede que $d_1 < d_2$ cuando ocurre E y $d_1 < d_2$ cuando no ocurre E, entonces, $d_1 < d_2$ (sin importar los eventos inciertos). Análogamente, si $d_1 \sim d_2$ cuando ocurre E y $d_1 \sim d_2$ cuando ocurre E, entonces $d_1 \sim d_2$.

En general, si dos opciones d_1 y d_2 tienen los mismos sucesos inciertos E_i y las consecuencias c_{1i} *dominan* a las de la segunda, de forma que $c_{1i} \geq_{2i} \ \forall i \in I$, entonces $d_1 \geq d_2$. Si además, $c_{1i} > c_{2i}$ para alguna $i \in I$, entonces $d_1 > d_2$.

4. **Eventos de referencia.** Este axioma establece que para poder tomar decisiones de forma razonable, es necesario medir la información del decisor expresándola en forma cuantitativa. Por lo tanto es necesario una medida ($\boldsymbol{P}$) basada en sucesos o eventos de referencia.

   - El decisor puede imaginar un procedimiento para generar puntos en el cuadro unitario de dos dimensiones, de manera tal que para cualquiera dos regiones R_1 y R_2 es ese cuadro, el evento $\{z \in R_1\}$ es más creíble que el evento $\{z \in R_2\}$ únicamente si el área de R_1 es mayor que el área de R_2.

5 Teoría de la Utilidad

- Hasta ahora, la discusión sobre construir decisiones Bayesianas ha utilizado el criterio de la consecuencia esperada, donde las consecuencias son medidas en utilidades monetarias.
- Sin embargo, las consecuencias en unidades monetarias pueden no reflejar la adecuadamente, la posición (preferencias) ante el riesgo de un decisor.

5.1 Evaluando la función de Utilidad

- La teoría de la utilidad se ha desarrollado tomando en cuenta la actitud del decisor hacia el riesgo.
- Desde el punto de vista Bayesiano, las decisiones se pueden ver como loterías las cuales incluyen probabilidades asociadas a los eventos inciertos y por lo tanto, asociadas a las posibles utilidades
- La teoría de utilidad hace algunos supuestos básicos (axiomas) sobre la consistencia de las preferencias del decisor entre loterías.

- Si estos axiomas son satisfechos, se puede mostrar que existe (**se puede construir**) una función de utilidad $U(C_{ij})$ que refleje adecuadamente la posición del decisor frente al riesgo.

5.2 Construcción de la función de Utilidad

El mayor problema al emplear la teoría de la utilidad es la construcción de la función de utilidad, por parte del decisor.

Ejemplo 5.1 *Un asesor financiero le propone un negocio especulativo a un cliente. Si el negocio es exitoso el cliente ganará \$10 mil, de otro modo ella perderá \$6 mil. Alternativamente, el cliente puede declinar el negocio, en tal caso su ganancia (consecuencia) es de \$0 -ni gan ni pierde. Así, existen tres posibles consecuencias.*

- ***Objetivo****: Asignar una utilidad (valores numéricos) a esas consecuencias (\$10, \$0 y \$-6) que reflejen la actitud (posición) del cliente ante el riesgo.*

5.3 Construcción de la función de Utilidad

- Desde que la utilidad es sinónimo de ganancia (subjetiva), entre más deseable sea una consecuencia mayor será la utilidad asignada a esta.
- No existe una única escala para la utilidad, por lo que se pueden definir dos puntos de la función de utilidad, de manera arbitraria (asociadas a dos consecuencias específicas).
- Una convención es utilizar: $u(c_*) = 0$ y $u(c^*) = 1$. Así, para nuestro ejemplo, $u(-6) = 0$ y $u(10) = 1$

5.3.1 Método de Referencia

- Existen varios métodos para evaluar una función de utilidad. Nosotros revisaremos, el denominado de *loterías de referencia*.
- Regresando al ejemplo del asesor financiero.
  - Definimos $u(-6) = 0$ y $u(10) = 1$. Supongamos que requerimos la utilidad de \$0, $u(0)$, así como de otras consecuencias, que nos permitan evaluar la forma de la función de utilidad del cliente.
  - Para encontrar $u(0)$, se le ofrece al decisor las siguientes loterías: una con ganancia segura de \$0 y la siguiente loteria de referencia:

Consecuencia	Probabilidad
10	0.5
-6	0.5

- Supongamos que el decisor prefiere ganar de manera segura \$0 a la lotería de referencia. Esto significa, que $u(0)$ excede la utilidad esperada de la lotería de referencia, la cual es de 0.5

$$u(10) * 0.5 + u(-6) * 0.5 = 1 * 0.5 + 0 * 0.5 = 0.5$$

- Desde que se desea encontrar un punto de indiferencia entre ganar de manera segura \$0 y una lotería de referencia, se necesita incrementar la probabilidad de ganar \$10 en la lotería de referencia.

- Supongamos que después de varios pasos, hemos encontrado la lotería de referencia

Consecuencia	Probabilidad
10	0.7
-6	0.3

- Es decir, en este momento el cliente indica que ella es indiferente entre \$0 seguros y esta lotería. Así, $u(0)$ será igual a la utilidad esperada de esta lotería de referencia, la cual es:

$$u(0) = u(10) * 0.7 * u(-6) * 0.3 = 1 * 0.7 + 0 * 0.3 = 0.7$$

- Para encontrar otros puntos de la función de utilidad, se procede de manera análoga. Por ejemplo, supongamos que se desea encontrar $u(2)$.

- Se le ofrece al cliente una serie de opciones entre ganar de manera segura \$2 y una lotería de referencia. Eventualmente, se alcanzará un punto de indiferencia, supongamos que esta es

Consecuencia	Probabilidad
10	0.8
-6	0.2

- Así,

$$u(0) = u(10) * 0.8 * u(-6) * 0.2 = 1 * 0.8 + 0 * 0.2 = 0.8$$

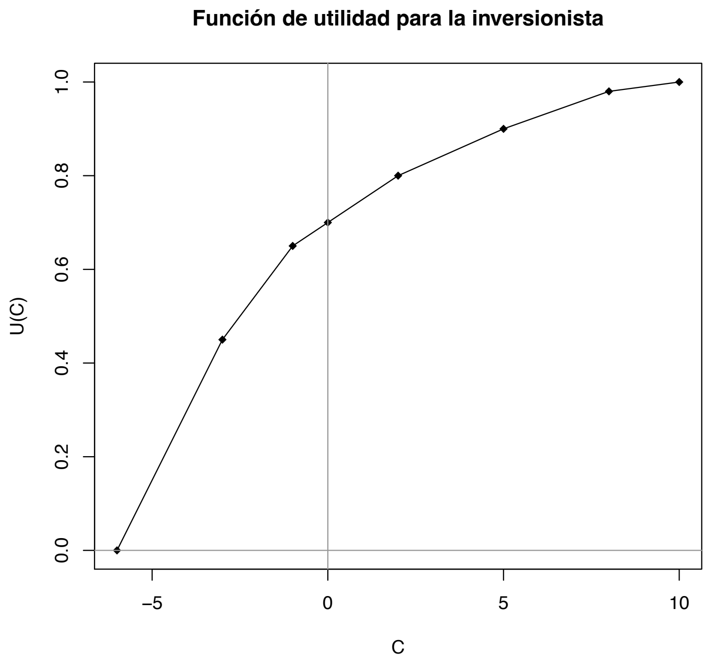
- Si se continua de esta manera, se puede obtener, por ejemplo, los siguientes resultados (utilidades) para las preferencias del cliente sobre la propuesta del negocio.

C	-6	-3	-1	0	2	5	8	10
$u(c)$	0	0.45	0.65	.70	0.80	0.90	0.98	1.0

- Cuando se obtienen suficientes puntos, se pueden graficar y a menudo se ajusta una función matemática a los datos, tales como:
  - Una función de utilidad cuadrática: $u(C) = b_0 + b_1C + b_2C^2$
  - Una función de utilidad exponencial: $u(C) = b_0 + b_1exp(b_2C^2)$

– Una función de utilidad lineal: $u(C) = b_0 + b_1 C$

- En nuestro ejemplo:



5.4 Comentarios

– La discusión anterior se centró en funciones de utilidad para el dinero. Sin embargo, se pueden desarrollar funciones de utilidad para otro tipo de consecuencias como: una tasa de retorno sobre algún negocio, periodos de recuperación de inversiones, localización del siguiente trabajo, etc.

- La utilidad es subjetiva. No se puede decir que un individuo esta o no equivocado en su actitud hacia el riesgo. Aún más, la función de utilidad de cualquier decisor puede cambiar en el tiempo cuando nuevas circunstancias ocurran. Finalmente, desde que la utilidad es subjetiva y en una escala arbitraria, tiene poco significado comparar funciones de utilidad para diferentes individuos.

- En muchas situaciones prácticas, la función de utilidad es aproximadamente lineal, por lo que las utilidades son directamente proporcionales a las consecuencias monetarias.

6 Implicaciones de los Axiomas de Coherencia

- Los axiomas de coherencia son la base de una teoría bien fundamentada, la Teoría de decisión Bayesiana.

- Una de las implicaciones de los axioma de coherencia es que, tanto las consecuencias como los sucesos inciertos pueden verse como casos particulares de opciones. Es decir:

- *Consecuencias:* $c \sim d_c = \{c|\Omega\}$, donde Ω es el evento seguro.

- *Eventos inciertos:* $E \sim d_E = \{c^*|E, c_*|E^c\}$, donde c^* y c_* son la "mejor" y la "peor" consecuencia.

6.1 Cuantificación de los eventos inciertos

Siguiendo la misma idea, si

$$d_E = \{c^*|E, c_*|E^c\},$$

existe un evento de referencia, R tal que

$$d_E \sim d_R = \{c^*|R, c_*|R^c\},$$

de, manera que E es igualmente creíble que R. Como la credibilidad de R se mide con su área, la credibilidad de E, también, es decir,

$$P(E) = Area(R)$$

Ejemplo 6.1 *¿Cómo asignar una probabilidad al evento A?.*

- *Idea:*

  – *Sean l_A y l_B, las loterías:*

$$l_A = \{c^*|A, c_*|\bar{A}\} \quad y \quad l_B = \{c^*|B, c_*|\bar{B}\},$$

  – *donde A y B son dos eventos inciertos. Entonces, se debe nota que comparar las loterías l_A y l_B es equivalente a comparar las verosimilitudes de los eventos A y B.*

Se consideran las siguientes loterías:

¿Cuál prefieres?:

- *Ganar c^* si ocurre A ó Ganar c_* si ocurre A^c.*

- *Ganar c^* con probabilidad p ó Ganar c_* con probabilidad $1-p$.*

$$d_A = \{c^*|A, c_*|A^c\} \quad ó \quad d_p = \{c^*|p, c_*|1-p\}$$

Por el axioma 1 es posible determinar si $d_A < d_p$, $d_A \sim d_p$ ó $d_A > d_p$.

La idea es encontrar el valor de p que haga que $d_A \sim d_p$. En este caso se satisface necesariamente que

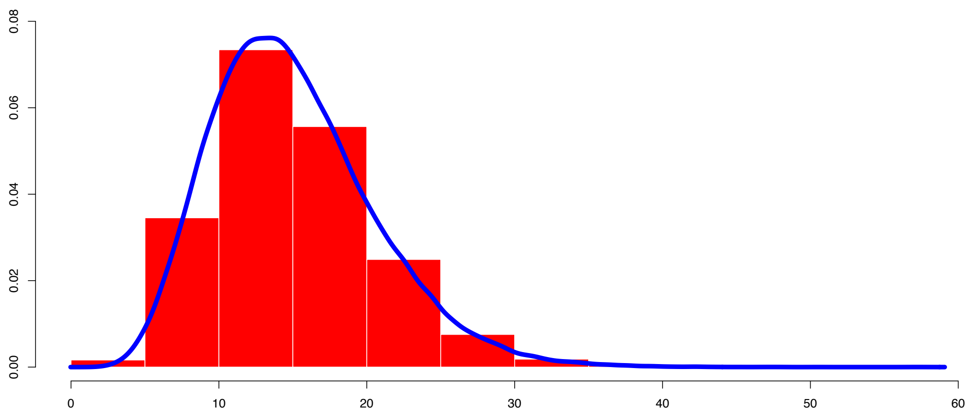
$$E\{u(d_A)\} = E\{u(d_p)\},$$

pero como

$$\begin{aligned} E\{u(d_A)\} &= u(c^*)P(A) + u(c_*)P(A^c) \\ &= (1)P(A) + (0)P(A^c) \\ &= P(A). \end{aligned}$$

Por lo tanto, $\qquad P(A) = p.$

Finalmente, se aplica este mismo procedimiento a cada uno de los eventos inciertos, digamos, $E_1 < E_2 < \ldots < E_k$. *Si el número de consecuencias es muy grande o incluso infinito la función de utilidad se puede aproximar por un modelo (discreto o continuo) obteniéndose de la siguiente forma,*



Modelo continuo . Si $E_\theta = \{\theta\} \quad \Rightarrow \quad \boldsymbol{E} = \{\theta | \theta \in [a, b]\}$

6.2 Cuantificación de las consecuencias

- Cuantificación de las consecuencias: Sabemos entonces que,

$$\begin{aligned} c_* \sim d_\emptyset = \{c^* | \emptyset, c_* | \Omega\} \\ c^* \sim d_\Omega = \{c^* | \Omega, c_* | \emptyset\} \end{aligned}$$

Si R_1 y R_2 son dos regiones, R_1 es más creíble que R_2 si

$$\text{Área}(R_1) > \text{Área}(R_2),$$

por lo tanto si

$$\begin{aligned} d_{R_1} = \{c^* | R_1, c_* | R_1^c\} \\ d_{R_2} = \{c^* | R_2, c_* | R_2^c\} \end{aligned}$$

Entonces, "graduando" R se tiene que para cualquier c tal que $c_* \leq c \leq c^*$ existe esa R tal que

$$c \sim d_R = \{c^*|R, c_*|R^c\}$$

Finalmente, una forma de cuantificar las consecuencias es tomando:

$$u(c) = \text{Área}(R)$$

NOTA: $u(c_*) = 0$ y $u(c^*) = 1$

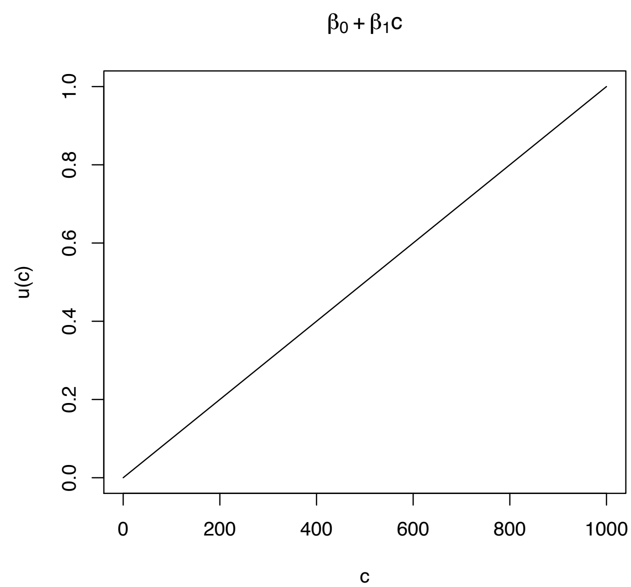
Ejemplo 6.2 ***Utilidad del dinero*** *Supongamos que la peor y la mejor consecuencia al jugar un juego de azar son:*

$$c_* = \$0 \quad (\textit{la peor})$$
$$c^* = \$1000 \quad (\textit{la mejor})$$

La idea es determinar una función de utilidad para cualquier consecuencia c tal que $c_* \leq c \leq c^*$.

1. *Una primera opción es asignar una función lineal: $u(c) = \beta_0 + \beta_1 c$.*

¿Será cierto que entre más dinero se tenga, la utilidad se incremente en forma lineal?



2. *Una segunda forma de asignar la utilidad es siguiendo el razonamiento sugerido por los axiomas: Sabemos que $u(c_*) = 0$ y $u(c^*) = 1$, se comparan las siguientes loterías:*

¿Cuál prefieres?:

- *Ganar seguro c.*
- *Ganar c^* con probabilidad p ó Ganar c_* con probabilidad $1-p$.*

$$d_c = \{c|\Omega\} \quad \text{ó} \quad d_p = \{c^*|p, c_*|1-p\}$$

Por el axioma 1 es posible determinar si $d_c < d_p$, $d_c \sim d_p$ *ó* $d_c > d_p$.

La idea es encontrar el valor de p *que haga que* $d_c \sim d_p$. *En este caso se satisface necesariamente que*

$$E\{u(d_c)\} = E\{u(d_p)\},$$

pero como

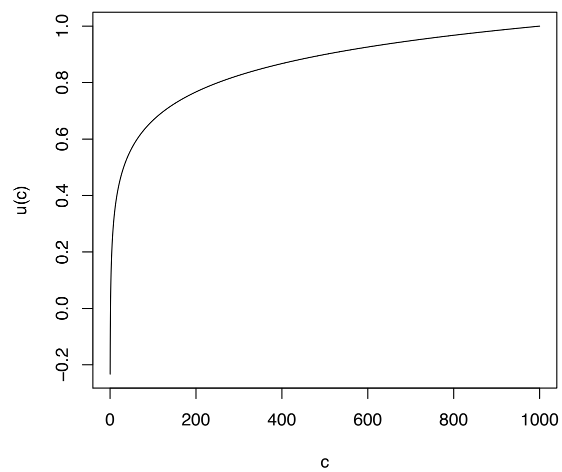
$$E\{u(d_c)\} = u(c),$$

porque c *es una consecuencia segura y*

$$\begin{aligned} E\{u(d_p)\} &= u(c^*)p + u(c_*)(1-p) \\ &= (1)p + (0)(1-p) \\ &= p. \end{aligned}$$

Por lo tanto, $u(c) = p$.

Finalmente, se aplica este mismo procedimiento a cada una de las consecuencias, digamos, $c_* < c_1 < c_2 < c_3 < c_4 < c^*$. *Si el número de consecuencias es muy grande o incluso infinito la función de utilidad se puede aproximar por un modelo obteniéndose de la siguiente forma,*



Entre más dinero se tiene el incremento en utilidad es menor

7 Principio de la utilidad esperada: Criterio Bayesiano

Hay que recordar que cualquier opción se puede escribir como

$$\begin{aligned} d &= \{c_1|E_1, c_2|E_2, ..., c_k|E_k\} \\ &\sim \{\{c^*|R_1, c_*|R_1\}|E_1, c_2|E_2, ..., c_k|E_k\} \\ &\sim \{c^*|R_1 \cap E_1, c_*|R_1^c \cap E_1, c_2|E_2, ..., c_k|E_k\} \\ &\vdots \\ &\sim \{c^*|B, c_*|B^c\} \end{aligned}$$

donde $B = (R_1 \cap E_1) \cup (R_2 \cap E_2) \cup ... \cup (R_k \cap E_k)$.

Si d_1 y d_2 son dos opciones, entonces existen B_1 y B_2 tales que

$$d_1 \sim \{c^*|B_1, c_*|B_1^c\}$$
$$d_2 \sim \{c^*|B_2, c_*|B_2^c\}$$

por lo tanto, $d_1 < d_2$ sólo si B_2 es más probable que B_1.
Finalmente, B_2 es más probable que B_1 sólo si

$$\sum u(c_{1i})P(E_i) < \sum u(c_{2i})P(E_i)$$

es decir,

$$E(u(d_1)) \quad < \quad E(u(d_2))$$
$$\Downarrow$$

Principio de la utilidad esperada máxima

Para un análisis más detallado, ver por ejemplo Capítulo 2 de Bernardo y Smith (1994)

8 Información inicial

- Como ya vimos anteriormente es necesario cuantificar los sucesos inciertos pertenecientes al espacio $\boldsymbol{E}$. A la cuantificación (inicial) de los eventos inciertos se le conoce como **información inicial** .

- Una forma de especificar la información inicial es siguiendo un procedimiento coherente dado por los axiomas mediante la comparación de opciones equivalentes y el uso de “loterías”.

- Otra manera de especificar la información inicial es asignando una distribución de probabilidad directamente sobre los eventos, de tal manera que ésta refleje el conocimiento inicial que se tenga.

- Consideremos el caso más sencillo: Supongamos que el espacio de eventos $\boldsymbol{E}$ es un conjunto discreto (posiblemente infinito), es decir, $E = \{E_1, ...E_m\}$. Sea θ una cantidad aleatoria que toma valores $1, 2, ..., m$ de tal manera que

$$P(E_i) = P(\theta = i)$$

  Entonces, en lugar de asignar una probabilidad directamente sobre $\boldsymbol{E}$, resulta más sencillo asignar una probabilidad sobre θ.

- La distribución de probabilidad sobre θ describe entonces la **información inicial** o *a priori* que se tiene inicialmente sobre el valor de θ. Esta distribución de probabilidad recibe el nombre de **distribución inicial** de θ.

- Como θ es una cantidad aleatoria discreta (en este caso), su distribución de probabilidad puede ser descrita mediante su función de densidad $f_\theta(i) = P(\theta = i) = p_i$, $i = 1, 2, ..., m$ y las probabilidades p_i pueden ser determinadas mediante relaciones entre ellas dadas por el decisor.

Ejemplo 8.1 ***Diagnóstico*** *Las consecuencias de un determinado tratamiento dependen de la enfermedad del paciente. Se considera que existen 5 enfermedades $\theta_1, \theta_2, \theta_3, \theta_4$ y θ_5 compatibles con los síntomas observados.*

Los médicos expertos determinan que existen las siguientes relaciones entre ellas

$$P(\theta_1 \cup \theta_2) = P(\theta_3 \cup \theta_4 \cup \theta_5)$$

$$P(\theta_2) = P(\theta_4) = 4P(\theta_3)$$

y creen muy remota la posibilidad de que se trate de la enfermedad θ_5. Determine la correspondiente distribución inicial.

Sea $p_i = P(\theta_i)$ $\forall i = 1, 2, 3, 4, 5$. Con la información dada por lo médicos se puede plantear el siguiente sistema de ecuaciones:

$$\begin{aligned} p_1 + p_2 &= p_3 + p_4 + p_5 \\ p_2 &= p_4 \\ p_4 &= 4p_3 \\ \sum p_i &= 1 \end{aligned}$$

Este sistema es un sistema con múltiples soluciones. Para resolver el sistema en forma sencilla, se pude tomar $p_5 = \gamma$. Así, el sistema tendrá la siguiente solución.

$$\begin{aligned} p_1 &= \tfrac{1+8\gamma}{10} \\ p_2 &= \tfrac{4(1-2\gamma)}{10} = p_4 \\ p_3 &= \tfrac{1-2\gamma}{10} \end{aligned}$$

En particular, si se juzga que es 20 veces más probable que θ_5 no se la enfermedad versus a que sí lo sea, entonces

$$\frac{1-\gamma}{\gamma} = 20 \;\Rightarrow\; \gamma = 0.048$$

Así, la distribución inicial que refleja el conocimiento inicial de los médicos está dada por:

$$\begin{aligned} p_1 &= 0.138 \\ p_2 &= 0.362 \\ p_3 &= 0.090 \\ p_4 &= 0.362 \\ p_5 &= 0.048 \end{aligned}$$

- Si el número de elementos de $\boldsymbol{E}$ es grande o incluso infinito, una forma de hacer la cuantificación de la información inicial es siguiendo el procedimiento dado por los axiomas para ciertos elementos de $\boldsymbol{E}$ y encontrar la cuantificación de los demás elementos ajustando un modelo continuo.

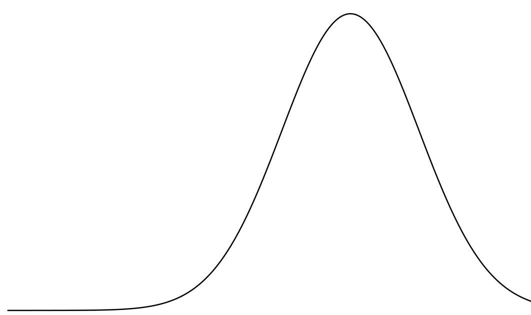
- Otra posibilidad para realizar la cuantificación de $\boldsymbol{E}$ (o de θ) es empezar directamente con un modelo probabilístico, averiguando primero algunas características cualitativas de la información que posee el investigador. Por ejemplo, se le puede preguntar:

  – ¿Cree que el modelo tiene una sola moda?

  – ¿Cree que el modelo es simétrico con respecto a esa moda?

  Si se tiene una respuesta afirmativa para ambas preguntas, se podría utilizar un modelo de la forma ...

¡ Normal !



Considerando algunas propiedades del modelo propuesto, y con la ayuda del "*tomador de decisiones*" se puede especificar el modelo completamente: $\mu =??$, $\sigma =??$.

Por ejemplo, se sabe que bajo un modelo Normal el 95% central de la probabilidad, se encuentra en el intervalo $(\mu - 2\sigma, \mu + 2\sigma)$

- ¿ **Dónde se encuentra la moda** ? $\Rightarrow \mu$

- ¿ **Entre que valores cree que se encuentre aprox. el 95% de la probabilidad** ? $\Rightarrow \sigma$

Ejemplo 8.2 ***Cantidad de tirosina*** *Las consecuencias de un determinado medicamento pueden determinarse a partir de la cantidad de tirosina contenida en la orina. La información inicial sobre la cantidad de tirosina θ contenida en la orina de un determinado paciente puede describirse de tal manera que se encuentre alrededor de 30 mg./24hrs. y que el porcentaje de veces que la cantidad de tirosina excede 49 mg./24hrs. sea de 25%.*

Determine la correspondiente distribución inicial.

- *De acuerdo con la información proporcionada, si se considera que una distribución Normal puede modelar adecuadamente la información inicial, entonces: $\theta \sim N(\mu, \sigma^2)$.*

  – *Cantidad de tirosina (θ) alrededor de 39* $\quad\Rightarrow\quad \mu = 39$

– $P(\theta > 49) = 0.25 \qquad \Rightarrow \qquad \sigma = 14.81$ *(¿Cómo?)*

$$
\begin{aligned}
P(\theta > 49) = P(Z > \frac{49-39}{\sigma}) = 0.25 \\
\Rightarrow Z_{0.25} = \frac{49-39}{\sigma}, \quad \textit{como} \;\; z_{0.25} = 0.675 \\
\Rightarrow \sigma = \frac{10}{0.675} = 14.81.
\end{aligned}
$$

Por lo tanto,

$$\theta \sim N(39, (14.81)^2).$$

- Una vez determinada la distribución inicial deben calcularse algunas probabilidades a partir de ella. Esto, por las siguientes razones:
  1. Para comprobar si las probabilidades calculadas son *consistentes* con la información inicial con la que se disponía.
  2. Corroborar si la familia elegida permite una buena *descripción* de la información inicial.

Ejemplo 8.3 ***Cantidad de tirosina (continuación ...)*** *El investigador opina que es muy poco probable que la cantidad de tirosina sea menor que 10 mg./24hrs. verificar que la distribución inicial es consistente con esta información.*

$$
\begin{aligned}
P(\theta < 10) &= P(Z < \frac{10-39}{14.81}) \\
&= P(Z < -1.95) \\
&= 0.256
\end{aligned}
$$

Por lo anterior, se puede decir que la distribución inicial es consistente con la información inicial.

9 Teorema de Bayes: El proceso de aprendizaje

Sean A y B dos eventos de un mismo espacio muestral Ω, entonces,

- **Probabilidad Condicional** :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

por otro lado, como

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \quad \Rightarrow \quad P(B \cap A) = P(B|A)P(A) = P(A \cap B)$$

Entonces:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \leftarrow \quad \textbf{Teorema de Bayes}$$

- **Independencia** : Se dice que dos eventos A y B son *independientes* sii:

$$P(A \cap B) = P(A)P(B)$$

- De la definición de probabilidad condicional, A y B son *independientes* sii:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)P(A)}{P(A)} = P(B)$$

  - **La ocurrencia de B no afecta la ocurrencia o no ocurrencia de A .**
  - **La ocurrencia de A no afecta la ocurrencia o no ocurrencia de B .**

9.1 El proceso de aprendizaje

- La reacción natural de cualquiera que tenga que tomar una decisión cuyas consecuencias dependan de la ocurrencia de eventos o sucesos inciertos ($\boldsymbol{E}$ ó θ), es intentar reducir su incertidumbre obteniendo más información sobre los eventos inciertos.
- Lo ideal sería recolectar o adquirir información que *elimine* por completo la incertidumbre, lo cual es muy caro o imposible.
- La idea es entonces recolectar información que *reduzca* la incertidumbre de los eventos inciertos, o equivalentemente, que mejore el conocimiento que se tiene sobre $\boldsymbol{E}$ ó θ.
- Esta información, generalmente, se basa en la información muestral que se obtiene de encuestas, estudios previos, experimentos, etc.
- El problema central de la inferencia estadística es el de proporcionar una metodología que permita *asimilar* la información accesible con el objeto de mejorar nuestro conocimiento inicial.
- Sea Z la información adicional que de alguna manera se pudo obtener sobre el evento E

¿ Cómo utilizar Z para mejorar el conocimiento sobre E ?

- Recordemos que la información que inicialmente se tiene sobre E se representa mediante una probabilidad, $P(E)$. El objetivo es **actualizar** la probabilidad de ocurrencia del mismo evento E dada la información disponible Z, i.e.,

$$P(E) \quad \overset{?}{\Rightarrow} \quad P(E|Z).$$

- La forma de hacer lo anterior es obteniendo la **probabilidad final** de $\boldsymbol{E}$, $P(E|Z)$.

$$P(E|Z) = \frac{P(Z|E)P(E)}{P(Z)}$$

donde

$$\begin{aligned}
E &: \text{ Evento incierto.} \\
Z &: \text{ Información adicional.} \\
P(E) &: \text{ Probabilidad inicial de } E. \\
P(Z|E) &: \text{ Probabilidad de } Z \text{ dado} E \\
P(Z) &: \text{ Probabilidad marginal de } Z. \\
P(E|Z) &: \text{ Probabilidad final de } E.
\end{aligned}$$

- Alternativamente, la probabilidad final de E se puede escribir como:

$$P(E|Z) \propto P(Z|E)P(E)$$

- En general, se tiene más de un suceso incierto sobre el cuál se quiere mejorar su conocimiento, esto da lugar al siguiente teorema.

Teorema de Bayes y Regla de Probabilidad Total : Sea $E_1, ...E_k$ una partición (eventos ajenos y exhaustivos) del espacio muestral Ω y se Z la información adicional sobre cada uno de los E_i, $i = 1, ..., k$. Entonces,

$$P(E_i|Z) = \frac{P(Z|E_i)P(E_i)}{P(Z)} \quad i = 1, 2, ...k$$

Como

$$1 = \sum_{i=1}^{k} P(E_i|Z) = \frac{\sum_{i=1}^{k} P(Z|E_i)P(E_i)}{P(Z)}$$

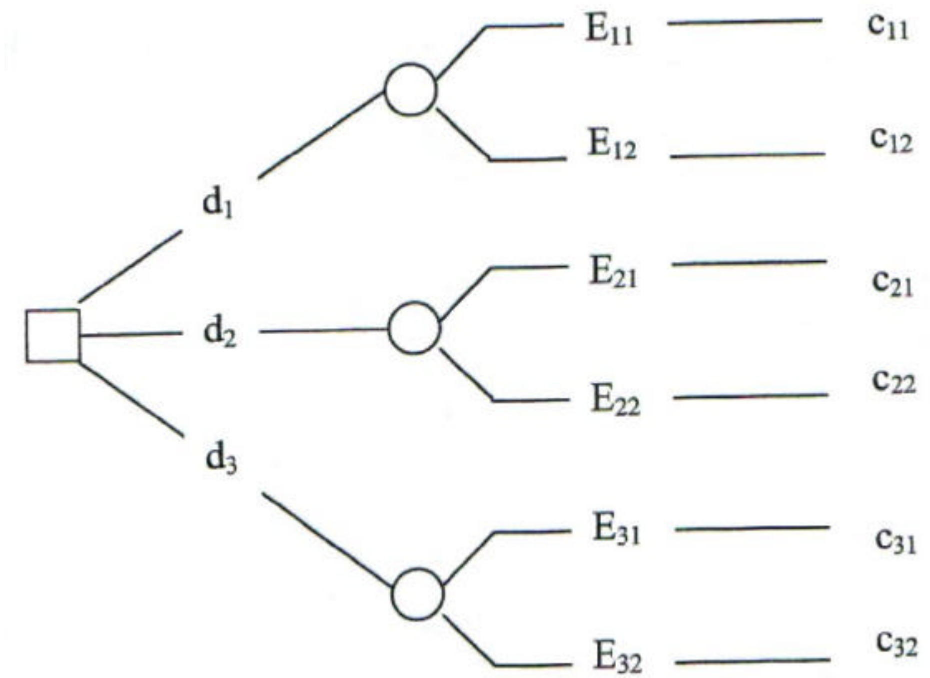
entonces,

$$P(Z) = \sum_{i=1}^{k} P(Z|E_i)P(E_i).$$

Finalmente,

$$P(E_i|Z) = \frac{P(Z|E_i)P(E_i)}{\sum_{i=1}^{k} P(Z|E_i)P(E_i)} \quad i = 1, 2, ...k$$

- ¿Para qué nos sirve reducir la incertidumbre de los eventos inciertos? Consideremos el siguiente problema de decisión:



- Se cuenta con lo siguiente:
  - $P(E_{ij})$: Cuantificación inicial de los eventos inciertos.
  - $u(c_{ij})$: Cuantificación de las consecuencias.
  - Z : Información adicional sobre los eventos inciertos.

Teorema de Bayes

$$P(E) \qquad \Rightarrow \qquad P(E|Z).$$

- Se tienen dos situaciones:
  1. Situación inicial o a-priori:

     $P(E_{ij})$, $u(c_{ij})$, $\quad \sum_j u(c_{ij})P(E_{ij}) \quad \leftarrow$ **Utilidad esperada inicial**
  2. Situación final o a-posterior:

     $P(E_{ij}|Z)$, $u(c_{ij})$, $\quad \sum_j u(c_{ij})P(E_{ij}|Z) \quad \leftarrow$ **Utilidad esperada final**

Ejemplo 9.1 *Suponga que una fábrica tiene 2 máquinas, A y B, que hacen el 60% y el 40%de la producción total respectivamente. La máquina A produce un 3% de artículos defectuosos, mientras que la máquina B produce un 5% de artículos defectuosos.*

a) ¿Cuál es la probabilidad de que un artículo producido, sea defectuoso?

b) Si se elige un artículo y este resulta defectuoso, ¿cuál es la probabilidad de que lo haya producido la máquina B?

- *Considere los eventos*

$$\begin{array}{lcl} A & \equiv & \textit{Artículo sea producido por la máquina } A \\ B & \equiv & \textit{Artículo sea producido por la máquina } B \\ D & \equiv & \textit{Artículo producido sea defectuoso} \end{array}$$

- *Entonces, de la información del problema se tiene:*

$$\begin{array}{lcl} P(A) & = & 0.6 \\ P(B) & = & 0.4 \\ P(D|A) & = & 0.03 \\ P(D|B) & = & 0.05 \end{array}$$

- *Así, la probabilidad (total) de que un artículo producido sea defectuoso es:*

$$P(D) = P(D|A)P(A) + P(D|B)P(B) = 0.38$$

- *Por otro lado, combinando la información inicial con la verosimilitud mediante el teorema de Bayes se tiene:*

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \frac{P(D|B)P(B)}{P(D|A)P(A) + P(D|B)P(B)} = 0.7105.$$

Así, es más probable (0.7105) que el artículo haya sido producido por la fábrica B dado que se ha observado que este es defectuoso. Se debe notar que $P(A|D) = 0.2895$.

Ejercicio 9.1 *Suponga que hay una prueba para detectar cáncer con la propiedad de que el 90% de aquellas personas con cáncer reaccionan positivamente y el 5% de aquellas sin cáncer reaccionan positivamente. Si el 1% de los pacientes en un hospital tienen cáncer, ¿cuál es la probabilidad de que un paciente seleccionado al azar, que reacciona en forma positiva a la prueba, realmente tenga cáncer?.*

Ejercicio 9.2 *Suponga que los coches tienen la misma probabilidad de ser fabricados en lunes, martes, miércoles, jueves o viernes. Los coches hechos en lunes tienen una probabilidad del 4% de ser amarillos; los coches hechos en martes, miércoles o jueves tienen una probabilidad del l% de ser amarillos; y los coches hechos en viernes tienen una probabilidad del 2% de ser amarillos.*

a) Si elige una coche al azar, del inventario global ¿Cuál es la probabilidad de que se amarillo?

b) Si se elige un coche al azar y resulta ser amarillo, ¿Cuál es la probabilidad de que haya sido fabricado en lunes?.

10 Actualización Secuencial

- ¿ Qué pasa si de alguna manera se obtiene aún más información adicional acerca del evento E ?

- ¿ Cómo se incorpora esta información **?**

- Existen dos caminos para actualizar la información que se obtiene sobre E:

  1. *Actualización secuencial.*
  2. *Actualización simultánea.*

Sea:
Z_1: información adicional obtenida sobre E.
Z_2: información adicional obtenida posteriormente sobre E.

10.1 Procedimiento secuencial

- Se cuenta con un conocimiento inicial $P(E)$.

- Actualizar el conocimiento para obtener $P(E|Z_1)$.

- Posteriormente, usar $P(E|Z_1)$ como conocimiento inicial y utilizar Z_2 para obtener $P(E|Z_1, Z_2)$.

¿Cómo se realiza esta actualización?

Paso 1: $P(E|Z_1) = \frac{P(Z_1|E)P(E)}{P(Z_1)}$

Paso 2: $P(E|Z_1, Z_2) = \frac{P(Z_2|Z_1,E)P(E|Z_1)}{P(Z_2|Z_1)}$

10.2 Actualización Simultánea

- Se cuenta con un conocimiento inicial $P(E)$.

- Usando Z_1 y Z_2 actualizar el conocimiento (simultáneamente) para obtener $P(E|Z_1, Z_2)$.

¿Cómo se realiza esta actualización?

Paso único: $P(E|Z_1, Z_2) = \frac{P(Z_1,Z_2|E)P(E)}{P(Z_1,Z_2)}$

- ¿Son equivalentes ambas formas de actualización?

$$
\begin{aligned}
P(E|Z_1, Z_2) &= \frac{P(Z_2|Z_1, E)P(E|Z_1)}{P(Z_2|Z_1)} \\
&= \frac{\{\frac{P(Z_1,Z_2,E)}{P(Z_1,E)}\}\{\frac{P(Z_1,E)}{P(Z_1)}\}}{\{\frac{P(Z_1,Z_2)}{P(Z_1)}\}} \\
&= \frac{P(Z_1, Z_2, E)}{P(Z_1, Z_2)} \\
&= \frac{P(Z_1, Z_2|E)P(E)}{P(Z_1, Z_2)}.
\end{aligned}
$$

Conclusión: **Ambas actualizaciones son equivalentes**

Ejercicio 10.1 *Supóngase que una caja contiene una moneda honesta y otra contiene una moneda con un sol en cada lado. Supóngase también que se selecciona una moneda al azar y que al lanzarla se obtiene un sol.*

a) Determine la probabilidad de que se haya seleccionado la moneda honesta. Verifique que esta probabilidad es $1/3$.

b) Supóngase que se lanza nuevamente la moneda y se obtiene otro sol. Existen dos maneras de determinar el nuevo valor de la probabilidad final de que la moneda sea honesta.

  – *La primera forma es volver al principio del experimento y suponer de nuevo que las probabilidades iniciales son* $P(M - honesta) = P(M - Nhonesta) = 1/2$. *Posteriormente, si se define el evento* S_1S_2 *como el suceso en que se obtiene sol en los dos lanzamientos de la moneda. Entonces, se debe calcular* $P(M - honesta|S_1S_2)$. *La cual se puede verificar que es* $1/5$.

- *La segunda forma de determinar esta misma probabilidad final es utilizar la información de que después de obtener un sol en el primer lanzamiento, la probabilidad final de $M-$ honesta es 1/3 y en consecuencia $P(M - Nhonesta) = 2/3$. Estas probabilidades finales pueden servir ahora como probabilidades iniciales para el siguiente paso del experimento. Por tanto, se puede suponer ahora que las probabilidades (iniciales) para $M - honestra$ y $M - Nhonesta$ son $P(M - honesta) = 1/3$ y $P(M - Nhonesta) = 2/3$, y calcular la probabilidad final $P(M - honesta|S_2)$. Se puede verificar que la probabilidad final del suceso $M - honesta$ obtenida de esta segunda forma es la misma que la obtenida en la primera forma.*

c) Finalmente, supóngase que se lanza nuevamente la moneda y se obtiene un águila como resultado. Obtenga la probabilidad (final) de que la moneda sea honesta. Antes de realizar cualquier cálculo, ¿cuál cree que debería ser este valor?.

Se puede hacer la siguiente afirmación general:

Si un experimento se lleva a cabo secuencialmente, entonces la probabilidad final de cualquier suceso también se puede calcular secuencialmente. Después de haber realizado cada paso, la probabilidad final del suceso encontrado en ese paso sirve como probabilidad inicial para el paso siguiente.

Ejercicio 10.2 *En algunos países, existe una ley que relaciona el derecho a conducir con el contenido de alcohol en la sangre. Una prueba rápida, realizada por la policía con un aparato portátil, tiene solamente una probabilidad de 0.8 de ser correcto entre las personas que la policía controla, esto es, de dar positivo si el contenido de alcohol en la sangre está por encima del límite legal y de 0.2 de dar negativo si se ha bebido más de lo permitido. Aquellos conductores a quienes la prueba les da positivo son sometidos al ministerio público a una prueba más rigurosa efectuada por un médico. Esta segunda prueba nunca da resultados incorrectos con un conductor sobrio, pero negativo en el 10% de los conductores detenidos que habían, de hecho, sobrepasado el límite legal. Si se asume que las pruebas son independientes.*

a) Determine la probabilidad que un conductor que ha sobrepasado el límite legal no sea sancionado, en función de la probabilidad p de que sea detenido por la policía cuando ha bebido.

b) Suponga que la probabilidad de ser detenido cuando ha bebido más de lo legal $p = 0.5$ y 0.1 en caso contrario (cuando no se ha bebido); y que el 15% de los conductores conducen con más alcohol en la sangre del permitido legalmente. Determinar la probabilidad de que un conductor haya bebido más de lo permitido en función de los resultados de las pruebas.

11 Procesos de inferencia como problemas de Decisión

11.1 Introducción

- El problema de inferencia paramétrico consiste en aproximar el verdadero valor de un parámetro θ dada una muestra de observaciones $X_1, ..., X_n$ provenientes de una población $f(x|\theta)$, donde $\theta \in \Theta$.
- El problema de inferencia sobre θ se puede plantear como un problema de decisión, donde:

  $\boldsymbol{D}$ = Decisiones de acuerdo al problema específico.

  $\boldsymbol{E} = \Theta$ (espacio parametral)

  $\boldsymbol{C} = \{(d, \theta) : d \in \boldsymbol{D}, \theta \in \Theta\}$

  $u(c) = u(d, \theta)$ = función de utilidad conveniente para cada problema ó

  $v(c) = v(d, \theta)$ = función de pérdida conveniente para cada problema

- Un punto importante es el de actualizar la información acerca de los eventos inciertos $\theta \in \Theta$.
- Por lo visto con los axiomas de coherencia, el decisor es capaz de cuantificar su conocimiento acerca de los eventos inciertos mediante una distribución de probabilidades.
- Definamos:
  - $f(\theta)$: **Distribución inicial** (o a-priori). Cuantifica el conocimiento inicial sobre θ.
  - $f(x|\theta)$: *proceso generador de información muestral*. Proporciona información adicional sobre θ.
  - $f(\boldsymbol{x}|\theta)$: **Función de verosimilitud** . Contiene toda la información sobre θ proporcionada por la muestra $\boldsymbol{x} = \{X_1, ..., X_n\}$.
- Toda la información sobre θ se combina para obtener un conocimiento final o a-posterior después de haber observado la muestra.

¿Cómo?

Mediante el *Teorema de Bayes*:

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)f(\theta)}{f(\boldsymbol{x})},$$

donde

$$f(\boldsymbol{x}) = \int_{\Theta} f(\boldsymbol{x}|\theta)f(\theta)d\theta \quad \text{ó} \quad f(\boldsymbol{x}) = \sum_{\theta} f(\boldsymbol{x}|\theta)f(\theta)$$

- Como $f(\theta|\boldsymbol{x})$ es una función de θ, una forma alternativa del teorema de Bayes es:

$$f(\theta|\boldsymbol{x}) \propto f(\boldsymbol{x}|\theta)f(\theta).$$

Distribución final $\propto$ Verosimilitud $\times$ Distribución inicial

Finalmente,

- $f(\theta|\boldsymbol{x})$: **Distribución final** (o a-posterior). Proporciona todo el conocimiento que se tiene sobre θ.

- NOTA: Al tomar θ el carácter de aleatorio, debido a que el conocimiento que tenemos sobre el verdadero valor de θ es incierto, entonces la función de densidad que genera observaciones con información relevante para θ realmente es una función de densidad condicional: $f(x|\theta)$.

- ¿Cómo se obtiene la función de verosimilitud?

  La función de verosimilitud es la función de densidad (condicional) conjunta de la muestra aleatoria vista como función del parámetro, i.e.,

$$f(\boldsymbol{x}|\theta) = f(x_1, ...x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

Distribución predictiva : La distribución predictiva es la función de densidad (marginal) $f(x)$ que permite determinar qué valores de la v.a. X resultan más probables.

- Lo que se conoce acerca de X esta condicionado al valor del parámetro θ, i.e. $f(x|\theta)$ (su función de densidad condicional). Como θ es un valor desconocido, $f(x|\theta)$ no puede utilizarse para describir el comportamiento de la v.a. X.
- Sin embargo, aunque el verdadero valor de θ es desconocido, siempre se dispone (de alguna manera) de cierta información sobre θ (mediante su distribución inicial $f(\theta)$. Esta información puede combinarse para poder dar información sobre los valores de X.

¿Cómo se obtiene la distribución predictiva?

- La distribución predictiva se obtiene de la siguiente manera:

$$f(x) = \int f(x|\theta)f(\theta)d\theta \quad \text{ó} \quad f(x) = \sum_{\theta} f(x|\theta)f(\theta)$$

- AHORA SUPONGAMOS que se cuenta con información adicional (información muestral) $x_1, ..., x_n$ de la densidad $f(x|\theta)$, por lo que es posible tener un conocimiento actualizado sobre θ mediante su distribución final $f(\theta|\boldsymbol{x})$.

- El objetivo es obtener información sobre los posibles valores que puede tomar una nueva variable X_* de la misma población $f(x|\theta)$. Si $f(x|\theta)$ es independiente de la muestra $\boldsymbol{x} = \{x_1, ...x_n\}$, entonces, la **distribución predictiva** final está dada por:

$$f(x_*|\boldsymbol{x}) = \int f(x_*|\theta)f(\theta|\boldsymbol{x})d\theta \quad \text{ó} \quad f(x_*|\boldsymbol{x}) = \sum_{\theta} f(x_*|\theta)f(\theta|\boldsymbol{x})$$

Ejercicio 11.1 *En cierta localidad el número de accidentes anuales de automóvil se puede modelar como una v.a.* $Y \sim Poisson(\lambda)$*, donde* λ *depende del conductor. Si se selecciona a un conductor aleatoriamente de esta localidad, se puede tener una variedad de valores de* λ*. Así, se puede representar la información inicial sobre este parámetro a través de una distribución de probabilidad* $f_\Lambda(\lambda|\cdot)$*, digamos* $Gamma(\lambda|\alpha_0, \beta_0)$*.*

a) Obtenga la distribución final del parámetro Λ*,* $f_\Lambda(\lambda|y)$*.*

b) Suponga que la inicial para el parámetro Λ *es una* $Gamma(10, 2)$*. Si se observa un registro, es decir un dato* $y = 12$*, ¿cómo se actualiza la información inicial sobre* Λ*?. Es decir, quién es la distribución final de* Λ*, dado* $y = 2$*.*

Respuesta: a)

$$\begin{aligned}
f_{\Lambda|Y}(\lambda|y) &= \frac{f_{Y|\Lambda}(y|\lambda)f_\Lambda(\lambda|\cdot)}{f_Y(y)} \\
&= \frac{f_{Y|\Lambda}(y|\lambda)f_\Lambda(\lambda|\cdot)}{\int_0^\infty f_{\Lambda,Y}(\lambda,y)d\lambda} \\
&= \frac{Poisson(y|\lambda)Gamma(\lambda|\alpha_0,\beta_0)}{\int_0^\infty Poisson(y|\lambda)Gamma(\lambda|\alpha_0,\beta_0)d\lambda} \\
&\propto Poisson(y|\lambda)Gamma(\lambda|\alpha_0, \beta_0) \\
&\propto \lambda^y exp\{-\lambda\}\lambda^{\alpha_0-1}exp\{-\beta_0\lambda\} \\
&= \lambda^{(\alpha_0+y)-1}exp\{-\lambda(\beta_0 + 1)\} \\
&\equiv kernel(Gamma(\alpha_0 + y, \beta_0 + 1)).
\end{aligned}$$

Por lo tanto, la distribución final del parámetro Λ *es una distribución Gamma con parámetros* $\alpha_F = \alpha_0 + y$ *y* $\beta_F = \beta_0 + 1$*.*

Respuesta: b)

Se debe notar que si se emplea una inicial $Gamma(10, 2)$*, para reflejar el conocimiento inicial sobre* λ*, en particular se está diciendo que se cree que el valor esperado (inicial) de* λ *es*

$$E(\Lambda) = \frac{\alpha_0}{\beta_0} = 5$$

Si se observa un dato y este es igual a $y = 12$ *accidentes anuales. Entonces, el conocimiento sobre* λ *se puede actualizar y la correspondiente distribución final resulta una* $Gamma(\alpha_F = 22, \beta_F = 3)$*.*

Así, en particular, la distribución final nos dice, que a la luz de los datos obtenidos, el valor esperado (final) para λ se actualiza y resulta ser

$$E(\Lambda|y=12) = \frac{\alpha_F}{\beta_F} = 7.33$$

Ejercicio 11.2 ***Cantidad de tirosina (continuación ...)*** *La cantidad de tirosina en la orina tiene una distribución inicial*

$$\theta \sim N(39, 219.47)$$

Para adquirir información sobre las cuales se encuentra un determinado paciente, se mide la cantidad de tirosina contenida en su orina. Debido a errores de medición, el valor obtenido no será en general el verdadero valor de θ, sino una variable aleatoria la cual se puede modelar con una distribución Normal centrada en θ y con desviación estándar de $\sigma = 2$ (que depende del aparato de medición). Es decir,

$$X|\theta \sim N(\theta, 4) \quad y \quad \theta \sim N(39, 219.47)$$

Se puede demostrar (tarea ...) que la distribución predictiva inicial toma la forma

$$X \sim N(39, 223.47)$$

¿Qué se puede derivar de esta distribución predictiva?

$$\begin{aligned} P(X > 60) &= P(Z > \tfrac{60-39}{\sqrt{223.47}}) \\ &= P(Z > 1.4047) = 0.0808. \end{aligned}$$

lo cual indica que es muy poco probable que una medición sea mayor a 60.

Con el objeto de mejorar la información inicial, se realizan 3 mediciones que resultan ser $x_1 = 40.62$, $x_2 = 41.8$ y $x_3 = 40.44$.

- *Se puede mostrar (* ***tarea*** *) que si $X|\theta \sim N(\theta, \sigma^2)$ y $\theta \sim N(\theta_0, \sigma_0^2)$ entonces, la distribución final de θ es una $N(\theta_1, \sigma_1^2)$, con*

$$\theta_1 = \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\theta_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0}} \quad y \quad \sigma_1^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}.$$

En nuestro ejemplo:

$$\bar{x} = 40.9533,\ \theta_0 = 39,\ \sigma^2 = 4,\ \sigma_0^2 = 219.47 \text{ y } n = 3$$

Por lo que,

$$\theta_1 = 40.9415,\ \sigma_1^2 = 1.3252$$

Así, la distribución final de θ resulta ser

$$\theta|\boldsymbol{x} \sim N(40.9415, 1.3252)$$

Observación : En el ejemplo anterior, la distribución inicial para θ era una Normal, la del modelo probabilístico $X|\theta$ también era una Normal y la distribución final fue otra Normal. Esto no siempre ocurre, pero nos da pauta para la siguiente definición.

11.2 Familias Conjugadas

Se dice que una familia de distribuciones de θ es **conjugada** respecto a un modelo probabilístico $f(x|\theta)$ si para cualquier distribución inicial perteneciente a tal familia, la distribución final correspondiente pertenece a la **misma familia** .

- Las familias conjugadas surgieron por la necesidad de tener familias de distribuciones iniciales que facilitaran el cálculo de la distribución final.
- Actualmente, con el desarrollo de los métodos computacionales, ya no es una necesidad trabajar con las familias conjugadas. Se puede usar cualquier familia que refleje de manera adecuada nuestro conocimiento inicial y mediante adecuados métodos computacionales es posible obtener una muy buena aproximación de la distribución final.

	Verosimiltud	*Inicial*	*Final*
Normal-Normal			
	$X|\mu \sim N(\mu, \sigma^2)$	$\mu \sim Normal$	$\mu|\boldsymbol{x} \sim Normal$
Normal-Gamma			
	$X|\tau \sim N(\mu, \frac{1}{\tau})$	$\tau \sim Gamma$	$\tau|\boldsymbol{x} \sim Gamma$
Gamma-Gamma			
	$X|\beta \sim G(\alpha, \beta)$	$\beta \sim Gamma$	$\beta|\boldsymbol{x} \sim Gamma$
Exponencial-Gamma			
	$X|\beta \sim Exp(\beta)$	$\beta \sim Gamma$	$\beta|\boldsymbol{x} \sim Gamma$
Beta-Binomial			
	$X|\theta \sim Bin(n, \theta)$	$\theta \sim Beta$	$\theta|\boldsymbol{x} \sim Beta$
Poisson-Gamma			
	$X|\lambda \sim Poisson(\lambda)$	$\lambda \sim Gamma$	$\lambda|\boldsymbol{x} \sim Gamma$

11.3 Problemas de Inferencia

- Los problemas de inferencia estadística se pueden agrupar, a grandes rasgos, en áreas como: La estimación puntual de parámetros, la estimación por intervalos o regiones y el problema de contrastar hipótesis.

- **Estimación Puntual** . El problema de estimación puntual visto como un problema de decisión se plantea de la siguiente manera:

  - $\boldsymbol{D} = \boldsymbol{E} = \Theta$.
  - $v(\tilde{\theta}, \theta)$ la pérdida de estimar el verdadero valor del parámetro de interés θ por medio de $\tilde{\theta}$

    Algunas funciones de pérdida que se pueden considerar son: f. de *pérdida cuadrática*, f. de *pérdida absoluta*, f. de *pérdida vecindad*.

1. Función de pérdida cuadrática:

$$v(\tilde{\theta}) = A(\tilde{\theta} - \theta)^2, \text{ donde } A > 0$$

   En este caso, la decisión óptima que minimiza la pérdida esperada es:

$$\tilde{\theta} = E(\theta)$$

   La mejor estimación de θ con pérdida cuadrática es la media de la distribución de θ al momento de producirse la estimación.

2. Función de pérdida absoluta:

$$v(\tilde{\theta}) = A|\tilde{\theta} - \theta|, \text{ donde } A > 0$$

   En este caso, la decisión óptima que minimiza la pérdida esperada es:

$$\tilde{\theta} = Mediana(\theta)$$

   La mejor estimación de θ con pérdida absoluta es la mediana de la distribución de θ al momento de producirse la estimación.

3. Función de pérdida vecindad:

$$v(\tilde{\theta}) = 1 - I_{B(\epsilon,\ \tilde{\theta})}\theta,$$

   donde $B(\epsilon,\ \tilde{\theta})$ es la *bola* abierta centrada en $\tilde{\theta}$ de radio ϵ.

   En este caso, la decisión óptima que minimiza la pérdida esperada cuando $\epsilon \to 0$ es:

$$\tilde{\theta} = Moda(\theta)$$

   La mejor estimación de θ con pérdida vecindad es la moda de la distribución de θ al momento de producirse la estimación.

Ejercicio 11.3 *Sean $X_1, ...X_n$ una muestra de v.a. de una población tal que $x_i|\theta \sim Bernoulli(\theta)$. Si la información inicial que se tiene se puede describir mediante una distribución Beta, i.e., $\theta \sim (\alpha, \beta)$. Es posible mostrar que la distribución final de θ es también una distribución Beta con parámetros*

$$\alpha + \sum_{i=1}^{n} x_i \quad y \quad \beta + n - \sum_{i=1}^{n} x_i.$$

Ahora, si el objetivo es estimar puntualmente el parámetro θ, entonces:

- Si se utiliza una función de pérdida cuadrática, se debe elegir como estimador a

$$\tilde{\theta} = E(\theta|\boldsymbol{x}) = \frac{\alpha + \sum x_i}{\alpha + \beta + n},$$

- Si se utiliza una función de pérdida vecindad, se debe elegir como estimador a

$$\tilde{\theta} = Moda(\theta|\boldsymbol{x}) = \frac{\alpha + \sum x_i - 1}{\alpha + \beta + n - 2},$$

11.4 Problema de Inferencia: Inferencia por Regiones

El problema de estimación por intervalos, visto como un problema de decisión se plantea de la siguiente manera:

- $\boldsymbol{D} = \{I : I \in \Theta\}$, donde I es algún intervalo de probabilidad $(1 - \alpha)$, es decir I cumple con
$$\int_I f(\theta) d\theta = 1 - \alpha.$$
- $\boldsymbol{E} = \Theta$.
- $v(I, \theta) = \|I\| - I_I(\theta)$ la pérdida de estimar el verdadero valor del parámetro de interés θ por medio de I.
- Esta función refleja la idea intuitiva que para un α dado es preferible reportar un intervalo de probabilidad I^* cuya longitud sea mínima. Por lo tanto,

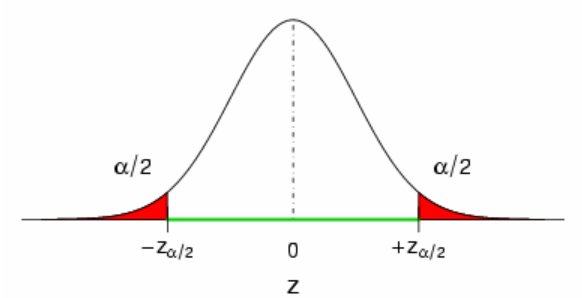
  La mejor estimación por intervalo de θ es el intervalo I^* cuya longitud sea mínima

Observación: El intervalo I^* de longitud mínima satisface la propiedad de ser un **intervalo de máxima densidad** , es decir, cumple con

$$\text{si } \theta_1 \in I^* \quad \text{y} \quad \theta_2 \notin I^* \quad \Rightarrow \quad f(\theta_1) \geq f(\theta_2)$$

- ¿Cómo se obtiene el intervalo de mínima longitud o de máxima densidad a-posterior?
  - Localizar la moda de la densidad final de θ.

– A partir de la moda trazar líneas rectas horizontales en forma descendiente hasta que se obtenga un intervalo donde se acumule $(1-\alpha)$ de probabilidad.



11.5 Problema de Inferencia: Contraste de Hipótesis

El problema de contraste de hipótesis es un problema relativamente sencillo y consiste en elegir entre k modelos probabilísticos definidos por k hipótesis H_0, H_1, etc. Aquí sólo revisaremos el caso $k = 2$.

Así, el problema de contraste de dos hipótesis visto como un problema de decisión se plantea de la siguiente manera:

- $\boldsymbol{D} = \boldsymbol{E} = \{H_0, H_1\}$.
- $v(d, \theta)$ la función de pérdida que toma la forma

$v(d,\theta)$	H_0	H_1
H_0	v_{00}	v_{01}
H_1	v_{10}	v_{11}

  donde v_{00} y v_{11} son las pérdidas de tomar una decisión correcta (generalmente $v_{00} = v_{11} = 0$), v_{10} es la pérdida de rechazar H_0 (aceptar H_1) cuando H_0 es cierta y v_{01} es la pérdida de no rechazar H_0 (aceptar H_0) cuando H_0 es falsa.

Las probabilidades iniciales son:

$$p_0 = P(H_0) \equiv \text{la probabilidad asociada a la hipótesis } H_0.$$
$$p_1 = 1 - p_0 = P(H_1) \equiv \text{la probabilidad asociada a la hipótesis } H_1.$$

Si $v_{00} = v_{11} = 0$, la pérdida esperada de aceptar cada una de las hipótesis es

$$\begin{aligned} E(v(H_0)) &= v_{01}(1-p_0) \\ E(v(H_1)) &= v_{10}p_0 \end{aligned}$$

Finalmente, la solución óptima es aquella que minimiza la pérdida esperada:

$$v_{01}(1-p_0) < v_{10}p_0 \quad \Leftrightarrow \quad \frac{p_0}{1-p_0} > \frac{v_{01}}{v_{10}} \quad \Rightarrow \boldsymbol{H_0}$$

$\boldsymbol{H_0}$ si p_0 es suficientemente grande comparada con $1-p_0$.

$$v_{01}(1-p_0) > v_{10}p_0 \quad \Leftrightarrow \quad \frac{p_0}{1-p_0} < \frac{v_{01}}{v_{10}} \quad \Rightarrow \boldsymbol{H_1}$$

$\boldsymbol{H_1}$ si p_0 es suficientemente grande comparada con $1-p_0$.

$$\frac{p_0}{1-p_0} = \frac{v_{01}}{v_{10}} \quad \Rightarrow \ \boldsymbol{H_0} \ \text{ó} \ \boldsymbol{H_1}$$

$\boldsymbol{H_0}$ ó $\boldsymbol{H_1}$ si p_0 no es suficientemente grande ni suficientemente pequeña comparada con $1-p_0$.

Si se cuenta con información adicional, se reemplaza $p_0 = P(H_0)$ por su probabilidad final $P(H_0|\boldsymbol{x})$. En este caso la decisión óptima sería

$$\frac{P(H_0|\boldsymbol{x})}{P(H_1|\boldsymbol{x})} > \frac{v_{01}}{v_{10}} \quad \Leftrightarrow \frac{P(\boldsymbol{x}|H_0)P(H_0)/P(\boldsymbol{x})}{P(\boldsymbol{x}|H_1)P(H_1)/P(\boldsymbol{x})} > \frac{v_{01}}{v_{10}} \frac{P(\boldsymbol{x}|H_0)}{P(\boldsymbol{x}|H_1)} > \frac{p_1 v_{01}}{p_0 v_{10}} \quad \Rightarrow H_0$$

H_0 si $P(H_0|\boldsymbol{x})$ es suficientemente grande comparada con $1-P(H_0|\boldsymbol{x})$.

$$\frac{P(H_0|\boldsymbol{x})}{P(H_1|\boldsymbol{x})} < \frac{v_{01}}{v_{10}} \quad \Leftrightarrow \frac{P(\boldsymbol{x}|H_0)P(H_0)/P(\boldsymbol{x})}{P(\boldsymbol{x}|H_1)P(H_1)/P(\boldsymbol{x})} < \frac{v_{01}}{v_{10}} \frac{P(\boldsymbol{x}|H_0)}{P(\boldsymbol{x}|H_1)} < \frac{p_1 v_{01}}{p_0 v_{10}} \quad \Rightarrow H_1$$

H_1 si $P(H_1|\boldsymbol{x})$ es suficientemente grande comparada con $1-P(H_1|\boldsymbol{x})$.

Ejercicio 11.4 *Sea $X|\mu \sim N(\mu,1)$ y se desean contrastar las siguientes hipótesis*

$$H_0 : \mu = 0 \quad v.s \quad H_1 : \mu = 1.$$

Si $p_0 = 2/3$, $v_{01} = 5$ y $v_{10} = 20$, entonces

$$\frac{p_0}{1-p_0} = \frac{2/3}{1/3} = 2 \quad y \quad \frac{v_{01}}{v_{10}} = \frac{5}{20} = 0.25$$

Por lo tanto, la decisión óptima es H_0 (no se rechaza H_0).

- *Sea $\boldsymbol{x} = (x_1, ..., x_n)$ información muestral, entonces*

$$P(\boldsymbol{x}|\mu) = f(\boldsymbol{x}|\mu) = \prod_{i=1}^{n} f(x_i|\mu)$$

por lo tanto

$$\begin{aligned} \frac{P(\boldsymbol{x}|\mu_0)}{P(\boldsymbol{x}|\mu_1)} &= \frac{(2\pi)^{n/2}\exp\{-\frac{1}{2}\sum(x_i-\bar{x})^2\}\exp\{-\frac{n}{2\sigma^2}\sum(\bar{x}-0)^2\}}{(2\pi)^{n/2}\exp\{-\frac{1}{2}\sum(x_i-\bar{x})^2\}\exp\{-\frac{n}{2\sigma^2}\sum(\bar{x}-1)^2\}} \\ &\vdots \\ &= \exp\{-\tfrac{n(2\bar{x}-1)}{2\sigma^2}\} \end{aligned}$$

De tal manera que se acepta H_1 (se rechaza H_0) si

$$\exp\{-\frac{n(2\bar{x}-1)}{2\sigma^2}\} < \frac{p_1 v_{01}}{p_0 v_{10}} \quad \Leftrightarrow \quad \bar{x} > \frac{1}{2} + \frac{2.0794}{n}.$$

12 ¿Fin?

- Modelación Bayesiana:
  - Modelos de regresión lineales y no lineales.
  - Modelos de Efectos Mixtos.
  - Modelos para datos longitudinales.
  - Modelos de Series de Tiempo.
  - Modelos multinivel.
  - Modelos de Diseño de Experimentos.
  - Técnicas de Simulación Estocástica.

$$\vdots$$

Fin ... ?

Agradecimientos

El apoyo del Departamento de Matemáticas de la UAM-I es agradecido ampliamente. El trabajo del primer autor también fue apoyado por el CONACyT, a través del fondo I0025 del programa de *Apoyos Complementarios para la Consolidación Institucional de Grupos de Investigación* (modalidad Repatriación).

13 Bibliografía

Referencias Básicas

- Bernardo. J. M. (1981). *Bioestadística. Una perspectiva Bayesiana.* Vincens-vives, Barcelona.
- Bernardo. J. M. y Smith, A. F. (1994). *Bayesian Theory.* John Wiley & Sons Ltd. Chichester, England.

Referencias Adicionales

1. Antelman, G. (1997). *Elementary Bayesian Statistics.* Editado por A. Madansky y McCulloch. Edward Elgar Publishing Ltd., Cheltenham.
2. Bernardo, J. M. (2003). *Bayesian Statistics.* Encyclopedia of Life Support Systems (EOLSS). Probability and Statistics. Oxford, UK: UNESCO
3. Carlin, B. P. y Louis, T. A. (2000) *Bayes and Empirical Bayes Methods for Data Analysis.* 2a. ed. Chapman & Hall, New York.
4. Congdon, P. (2001). *Bayesian Statistical Modelling.* John Wiley & Sons, Ltd. Chichester, England.
5. DeGroot, M. H. (1975). *Probability and Statistics.* 2a. ed. Addison-Wesley, Massachusetts, EU.
6. Gelman, A., Carlin, J. B., Stern, H. S. y Rubin, D. B. (2004). *Bayesian Data Analysis.* 2a. ed. Chpman & Hall, New York.
7. Lee P. M. (1997). *Bayesian Statistics. An introduction.* 2a. ed. Arnold, London y Jhon Wiley & Sons Inc., New York.
8. Lindley, D. V. y Philips, L. D. (1976). Inference for a Bernoulli process (a Bayesian view). Amer. Statist, 30, 112-119.
9. Mendoza, M. y Rugeiro, P. (2011). *Estadística Bayesiana.* Ed. Departamento de Estadística, ITAM. México, DF.
10. Nuñez-Antonio, G. y López-García A. (1995). Introducción a los modelos lineales Bayesianos. Tesis de licenciatura, Facultad de Ciencias, UNAM, México DF.
11. Nieto-Barajas, L. E. *Estadística Bayesiana.* Notas de clase. Ed. ITAM, México DF.

14 Ejercicios

14.1 Inferencias para una proporción poblacional: El modelo Beta-Binomial

1. Para determinar la efectividad de un periódico para alcanzar a su audiencia potencial, una compañía de investigación de mercados selecciona una muestra aleatoria de la audiencia potencial y los entrevista. De las 150 personas en la muestra, 29 han visto el último número.

   (a) ¿Cuál es la distribución de y, el número de personas que ha visto el último número?

   (b) Si se utiliza una distribución uniforme como inicial para p, la proporción audiencia potencial que ha visto el último número, ¿cuál es la distribución final de p?

2. En cierta ciudad se esta considerando construir un nuevo museo. Los periódicos locales desean determinar el nivel de apoyo de este proyecto, y realizarán una encuesta entre los residentes de la ciudad. De la muestra de 120 personas encuestadas, 74 apoyaron la construcción del museo.

   (a) ¿Cuál es la distribución de y, el número de personas que apoya la construcción del museo?

   (b) Si se utiliza una distribución uniforme como inicial para p, la proporción de la población que apoya la construcción del museo, ¿cuál es la distribución final de p?

3. Sophie, la editora del periódico de estudiantes, va a realizar una encuesta de estudiantes para determinar el nivel de apoyo para el actual presidente de la asociación de estudiantes. Ella necesita determinar su distribución inicial para p, la proporción de estudiantes quienes apoyan al presidente. Ella decide que su media inicial es .0.5, y que su desviación estándar es 0.15

   (a) Determine la densidad $beta(\alpha, \beta)$ que es acorde con el conocimiento inicial de Sophie.

   (b) ¿Cuál es el tamaño de muestra equivalente de la inicial de Sophie?

4. Sea p la proporción de estudiantes de una universidad que aprueban las políticas gubernamentales sobre becas estudiantiles. El periódico estudiantil toma una m.a. de $n = 30$ estudiantes de la universidad y registra si los estudiantes aprueban las políticas gubernamentales sobre asignación de becas.

   (a) ¿Cuál es la distribución de y, el número de respuestas "si"?

   (b) Suponga que de los 30 estudiantes, 8 contestaron que si. ¿Cuál es el estimador *frecuentista* de p?.

   (c) Encuentre la distribución final de p, $f(p|y)$, si se usa una inicial uniforme.

   (d) Cuál sería el estimador Bayesiano de p?.

5. El método estándar para la detección de una enfermedad falla para detectar la presencia de la enfermedad en 15% de los pacientes que en realidad tienen la enfermedad. Un nuevo método de detección para la presencia de la enfermedad se desarrollo. Una muestra aleatoria de $n = 75$

pacientes que se sabe que tienen la enfermedad es analizada utilizando el nuevo método. Sea p la probabilidad de que el nuevo método falle para detectar la enfermedad.

(a) ¿Cuál es la distribución de y, el número de veces que el nuevo método falla para detectar la enfermedad?

(b) De los 75 pacientes, el nuevo método falla en detectar la enfermedad en $y = 6$ casos. ¿Cuál es el estimador frecuentista de un p?

(c) Utilizando una inicial $Beta(1, 6)$, encuentra la distribución final de p, $f(p|y)$.

(d) Encuentre la media y la varianza posterior.

(e) Si $p \geq 0.15$, entonces el nuevo método de diagnóstico no es mejor que el método estándar. Contrasta, de manera Bayesiana

$$H_0 : p \geq 0.15 \quad vs. \quad H_1 : p < 0.15$$

con un nivel de significancia del 5%.

6. En el estudio de la calidad del agua en los arroyos de Nueva Zelanda documentados en McBride *et al.* (2002) un alto nivel de Campylobacter se define como un nivel mayor que 100 por cada 100 ml de agua corriente. $n = 116$ muestras fueron tomadas de arroyos que tiene un alto impacto ambiental para las aves. De estas $y = 11$ resultaron tener alto nivel de Campylobacter. Sea p la verdadera probabilidad de que una muestra de agua de este tipo de arroyos tenga un alto nivel de Campylobacter.

(a) Obtenga el estimador frecuentista de p.

(b) Utilice un $Beta(1, 10)$ como inicial y obtenga la distribución posterior de p.

(c) Encuentre la media y la varianza posterior. ¿Cuál es el estimador Bayesiano para p?

(d) Encuentre un intervalo de credibilidad del 95% para p.

(e) Contraste las hipótesis

$$H_0 : p = 0.10 \quad vs. \quad H_1 : p \neq 0.10$$

con un nivel de significancia del 5%.

7. En el mismo estudio de la calidad del agua $n = 145$ muestras fueron tomadas de arroyos que tiene un alto impacto ambiental para la producción láctea. De estas $y = 9$ resultaron tener alto nivel de Campylobacter. Sea p la verdadera probabilidad de que una muestra de agua de este tipo de arroyos tenga un alto nivel de Campylobacter.

(a) Obtenga el estimador frecuentista de p.

(b) Utilice un $Beta(1, 10)$ como inicial y obtenga la distribución posterior de p.

(c) Encuentre la media y la varianza posterior. ¿Cuál es el estimador Bayesiano para p?

(d) Encuentre un intervalo de credibilidad del 95% para p.

(e) Contraste las hipótesis

$$H_0 : p = 0.09 \quad vs. \quad H_1 : p \neq 0.09$$

con un nivel de significancia del 5%.

14.2 Inferencias para una media poblacional: El modelo Normal-Normal

8. El proceso estándar para producir un polímero tiene una media de 35%. Un ingeniero químico ha desarrollado un proceso modificado. El "corre" el proceso en 10 lotes y mide la producción (en porcentaje) para cada lote. Los datos son

38.7	40.4	37.2	36.6	35.9
34.7	37.6	35.1	37.5	35.6

Si se considera que la *producción* se puede describir con una distribución $Normal(\mu, 3^2)$.

(a) Use una inicial $Normal(30, 10^2)$ para μ. Obtenga la correspondiente distribución final.

(b) El ingeniero quiere conocer si el proceso modificado incrementa la producción media. Plantee el problema como un contraste de hipótesis estableciendo claramente la hipótesis nula y la hipótesis alternativa.

(c) Realice el contraste anterior con un 5% de significancia.

9. Un ingeniero toma una muestra de 5 vigas de acero de un lote, y mide la cantidad que se pandean bajo una carga estándar. Las cantidades en mm son:

5.19	4.72	4.81	4.87	4.88

Si sabe que el pandeo es $Normal(\mu, 0.25^2)$.

(a) Utilice una $Normal(5, 0.52)$ como inicial para μ. Encuentra la distribución posterior.

(b) Para que un lote de vigas sea aceptable, la media del pandeo bajo la carga estándar debe ser inferior a 5.20. ($\mu < 5.20$). Establezca el problema como un contraste de hipótesis especificando claramente la hipótesis nula y la hipótesis alternativa.

(c) Realice el contraste anterior con un 5% de significancia.

10. Un estadístico compra un paquete de 10 pelotas nuevas de golf, deja caer cada pelota a una altura de un metro, y registra las medidas de la altura del primer rebote en centímetros. Los diez valores son:

79.9	80.0	78.9	78.5	75.6	80.5	82.5	80.1	81.6	76.7

Se asume que y, la altura de rebote (en cm) es $Normal(\mu, \sigma^2)$, donde la desviación estándar $\sigma = 2$.

(a) Suponga una inicial $Normal(75, 10^2)$ y encuentre la distribución posterior de μ.

(b) Calcular un intervalo de credibilidad Bayesiano del 95% para μ.

(c) Realice el contraste de hipótesis bayesiano

$$H_0 : \mu \geq 80 \quad vs. \quad H_1 : \mu < 80$$

con un nivel de significancia del 5%.

11. El estadístico, del problema anterior, compra un paquete de 10 pelotas usadas que han sido recubiertas con un material líquido. El deja caer nuevamente cada pelota a una altura de un metro, y registra las medidas de la altura del primer rebote en centímetros. Los diez valores son:

73.1	71.2	69.8	76.7	75.3	68.0	69.2	73.4	74.0	78.2

Se asume que y, la altura de rebote (en cm) es $Normal(\mu, \sigma^2)$, donde la desviación estándar $\sigma = 2$.

(a) Suponga una inicial $Normal(75, 10^2)$ y encuentre la distribución posterior de μ.

(b) Calcular un intervalo de credibilidad Bayesiano del 95% para μ.

(c) Realice el contraste de hipótesis bayesiano

$$H_0 : \mu \geq 80 \quad vs. \quad H_1 : \mu < 80$$

con un nivel de significancia del 5%.

14.3 Problemas varios

12. Sea θ la probabilidad de recibir una respuesta incorrecta a una pregunta específica hecha por los agentes de la policía judicial al reportar un incidente telefónicamente. Suponga que la distribución inicial para θ es una distribución Beta con parámetros $\alpha = 3$ y $\beta = 10$. Si una muestra de 20 llamadas da como resultado $(0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0)$, es decir, 5 respuestas fueron incorrectas.

(a) ¿Cuál es la distribución posterior de θ ?

(b) Encuentra la media, la varianza y la moda de esta distribución posterior.

(c) Grafica la distribución posterior.

(d) Encuentra la probabilidad posterior de que $\theta < 0.3$.

(e) Encuentre un intervalo posterior de máxima densidad al 99%.

13. Suponga que las ventas de un determinado producto en un determinado estante se comportan como un modelo Poisson con intensidad 10 por semana. El producto es movido a un estante menos concurrido y estamos interesados en la nueva tasa de venta λ. Suponga que la mejor creencia que se tiene sobre el valor esperado de λ es 9.5 por semana con una desviación estándar para λ de 1.0 por semana.

    (a) Encuentre una distribución inicial Gamma para λ con las especificaciones que se tienen.

    (b) Para esta distribución inicial encuentra $P(\lambda < 10)$. En una muestra de 8 semanas de ventas para este producto dio como resultado (13,8,9,6,8,7,7,7).

    (c) Encuentra la distribución posterior para λ.

    (d) Encuentra la media y la varianza de esta distribución posterior.

    (e) Encuentra un intervalo posterior de máxima densidad al 95% para λ.

14. Un modelo Normal tiene varianza $\sigma^2 = 25$. Suponga que la distribución inicial para μ es Normal(100,4).

    (a) Encuentre la probabilidad inicial de que $\mu < 102$.

    (b) Encuentra un intervalo inicial de máxima densidad al 95% para μ.

    Una muestra de 50 observaciones del modelo Normal dio como resultado (105, 102, 102, 101, 121, 108, 108, 105, 102, 111, 104, 112, 112, 101, 98, 105, 100, 102, 101, 96, 110, 110, 108, 100, 105, 96, 95, 108, 97, 103, 104, 103, 98, 100, 102, 103, 109, 101, 106, 105, 108, 108, 104, 106, 109, 107, 100, 101, 106, 103)

    (c) Encuentra la probabilidad posterior de que $\mu < 102$.

    (d) Encuentra un intervalo posterior de máxima densidad al 95% para μ.

15. En un estudio dietético se determina el contenido proteínico en gramos de proteínas por 100gr. de materia comestible. El contenido proteínico de 50 porciones de frijoles arrojó un contenido promedio de 22 con una desviación estándar muestral de 5. Si se consideran observaciones Normales con media μ y varianza $\theta^2 = 26$, contrastar la hipótesis de que el valor medio supera los 23 grs. Considere que la función de utilidad es de la forma $u(H_i, H_j) = 1$ si $i = j$ y cero en caso contrario. Suponga además que la información inicial para μ se puede representar adecuadamente mediante una distribución $N(24, 100)$.

    antigripal. Sea θ la probabilidad de que el nuevo antigripal sea eficaz. Las hipótesis a contrastar son:

    $$H_0 : \theta > 0.95 \;\; contra \;\; H_1 : \theta \leq 0.95.$$

    Si se acepta H_0 siendo cierta, se obtiene un beneficio de 10 millones debido a la comercialización de un producto útil. Por otro lado, si se acepta siendo falsa, se incurre en una pérdida de 7 millones debido a que el producto ya comercializado debe ser retirado del mercado. Si se rechaza

H_0 el producto no es comercializado y se pierden 0.5 millones invertidos en investigación. Si la información inicial sobre θ se puede describir mediante una distribución Beta(15,3) y una prueba del nuevo antigripal en 50 pacientes arroja una reacción positiva en 47 de ellos, determine la decisión óptima suponiendo que la pérdida es proporcional al dinero.