UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

APLICACIONES DE PROCESOS DE DECISION MARKOVIANOS A PROBLEMAS DE SEGUROS

T E S I S

QUE PARA OBTENER EL TITULO DE

A C T U A R I O

P R E S E N T A :

JOSE CARLOS GONZALEZ RODRIGUEZ

DIRECTOR DE TESIS: DRA. GUADALUPE CARRASCO LICEA

FACULTAD DE CIENCIAS
UNAM

2004

FACULTAD DE CIENCIAS
SECCION ESCOLAR

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

Aplicaciones de procesos de decisión Markovianos a problemas de seguros

realizado por José Carlos González Rodríguez

con número de cuenta 9354739-0 , quien cubrió los créditos de la carrera de:

Actuaría.

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario Dra. Guadalupe Carrasco Licea

Propietario Dr. Juan González Hernández

Propietario Act. Rubén Ugalde Franco

Suplente Act. Marisa Miranda Tirado

Suplente M. en C. Hugo Villaseñor Hernández

Consejo Departamental de Matemáticas

Act. Jaime Vázquez Alamilla

FACULTAD DE CIENCIAS
CONSEJO DEPARTAMENTAL
DE
MATEMATICAS

Contenido

Capítulo 1

Dos problemas de seguros

1.1 Introducción

El objetivo de este capítulo es hacer una primera presentación de los dos problemas de seguros cuya resolución es el objetivo central de este trabajo. El primero de ellos consiste en establecer la política óptima de reclamos para un asegurado que ha contratado un seguro multiperíodo en el cual el monto de la prima se ve afectado por los reclamos realizados en los distintos períodos. El segundo se refiere a la determinación de primas óptimas por parte de una compañía aseguradora que desconoce la distribución precisa de reclamos. Se trata entonces de un problema planteado desde la óptica del comprador de un seguro y otro planteado desde el punto de vista de la compañía aseguradora.

Antes de presentar las hipótesis y las condiciones de cada uno de los modelos, en la sección 1.2 haremos una presentación informal de los procesos de decisión Markovianos que será la herramienta básica que usaremos en la resolución de los problemas planteados. En el siguiente capítulo de este trabajo formalizaremos los conceptos relacionados con procesos de decisión Markovianos y estudiaremos formas de obtener políticas óptimas.

Las secciones 1.3 y 1.4 están dedicadas a exponer los elementos de los modelos correspondientes a cada uno de los problemas, mismos que serán resueltos más adelante.

1.2 Los procesos de decisión Markovianos

Supongamos que una persona interactúa con un sistema dinámico cuya evolución está afectada por la presencia de uno o más elementos aleatorios. Esta persona está en posibilidades de tomar ciertas decisiones que afectan el desarrollo futuro del sistema aun cuando no lo determinan completamente por el carácter estocástico de su evolución.

El tomador de decisiones observa el estado x en que se encuentra el sistema en un momento t y sólo con base en esa observación obtiene toda la información necesaria para elegir una acción o control a. Como resultado de la decisión tomada, se genera una recompensa (o un costo) y el sistema evoluciona hacia un nuevo estado de acuerdo a una ley de probabilidad $p_t(\cdot \mid x, a)$ condicionada por el estado en que se encontraba y por la acción elegida. Tras un período de tiempo determinado, el tomador de decisiones vuelve a observar el estado y a elegir una acción. Este procedimiento se repite a lo largo de un lapso de tiempo, que puede ser finito o infinito, generando una sucesión de recompensas $\{r_t(x, a)\}$. A un sistema que tenga este tipo de evolución se le llama *sistema dinámico controlado a tiempo discreto.* El modelo matemático de un sistema así consta de cinco elementos:

- Un conjunto T de épocas de decisión que son los momentos en que se observa el estado del sistema.
- Un conjunto X de estados en los que puede encontrarse el sistema.
- Un conjunto A de acciones o controles que pueden ser elegidas por el tomador de decisiones
- Una probabilidad de transición $p_t(\cdot \mid x, a)$ que refleja el carácter aleatorio de la evolución del sistema.
- Una función recompensa $r_t(x, a)$ que representará un costo cuando sea negativa.

Antes de que el sistema empiece a evolucionar, los estados y las acciones son variables aleatorias cuya distribución depende de la distribución de probabilidades de los elementos aleatorios del sistema. El problema que se trata de resolver es cómo determinar *a priori* una política, es decir, una sucesión de acciones, que conduzcan al mejor desarrollo posible del sistema. Para ello, se requiere establecer un criterio de optimalidad dado por una función de los valores esperados de las recompensas a la que llamaremos *índice de funcionamiento.* El valor de este índice nos permitirá comparar el comportamiento del sistema para distintas acciones seleccionadas.

Se requiere también conocer las distintas clases de políticas que se pueden aplicar. Una política está formada por una colección de funciones $d_1, d_2, ...d_t, \ldots$ que indican la forma de elegir una acción en cada momento t. Cada una de las funciones d_t se conoce como *función de decisión al tiempo t.*

Modelos como éste han sido utilizados en el análisis de una gran cantidad de situaciones en distintas áreas del conocimiento.Veamos un par de ejemplos.

Ejemplo 1.2.1 ***Administración de un inventario.*** *El problema del control de inventarios consiste en general en determinar cuántos productos se deben tener a la mano para lograr satisfacer la demanda aleatoria sin incurrir en gastos innecesarios por el almacenamiento y el manejo de los productos. El administrador observa la cantidad de productos que tiene a la mano al principio de cada día o de cada semana o de cada período que se requiera, y decide en qué momento ordenar la compra de más productos a la unidad de producción (punto de reorden) y cuántos productos se deben adquirir (nivel de reordenamiento). Así, se describe un problema secuencial para determinar puntos y niveles óptimos de reordenamiento. Supongamos que se trabaja un solo tipo de productos y que cada producto es una unidad indivisible. En el caso más sencillo, se considera que los clientes a los que no se les pueda surtir su demanda en el momento que la solicitan, se pierden; es decir, no se surten pedidos atrasados.*

En este caso, las épocas de decisión son los momentos de revisión del inventario a la mano. El estado en que se encuentra el sistema es la cantidad de productos almacenados al momento de la observación. Las acciones que el administrador puede tomar corresponden a las distintas cantidades de productos que puede ordenar incluyendo la decisión de no adquirir productos. La probabilidad de transición depende de la distribución de la demanda aleatoria a lo largo del período entre dos épocas de decisión.

La función recompensa debe contemplar los ingresos obtenidos por las ventas y los costos asociados al manejo y almacenamiento de productos, así como los gastos por ordenar la compra de productos adicionales. A veces también se incluye una penalización por no satisfacer la demanda debido a que esto conduce a la pérdida de clientes. Si el proceso de revisiones periódicas termina en algún momento $t = N$, entonces en ese momento ya no se toma ninguna decisión, simplemente se hace un recuento de los artículos que quedan en el almacén y se considera una recompensa terminal igual al valor de esos productos.

El índice de funcionamiento puede ser la suma de las recompensas totales esperadas desde $t = 1$ hasta $t = N$, o bien la suma de los valores presentes de las recompensas esperadas. Evidentemente, el objetivo es encontrar una política que maximice cualquiera de estos índices de funcionamiento.

Una regla de decisión (determinista) especifica la cantidad a ordenar al tiempo t como función de la cantidad de productos que se tienen a la mano, y una política consiste de la sucesión de estas funciones.

Si no hay restricciones adicionales, la política óptima tiene una forma muy sencilla que ha demostrado su efectividad en la práctica: hacer un pedido sólo hasta el momento en que la cantidad de productos a la mano sea menor o igual que un límite inferior establecido de antemano. Cuando esto suceda, la política óptima sugiere ordenar la cantidad necesaria para llegar hasta un límite superior también preestablecido. El

problema entonces se reduce a encontrar los valores de estos dos límites para cada caso particular.

Las condiciones que deben tomarse en cuenta son: la capacidad M del almacén, el costo y el precio de venta del artículo que supondremos fijos a lo largo de todos los períodos de evolución del sistema, y la distribución de probabilidades de la demanda aleatoria.

Partiendo de estas condiciones, podemos establecer los elementos del modelo de la siguiente forma:

El conjunto de épocas de decisión es

$$T = \{1, 2, \ldots N\}.$$

Los espacios de estados y de acciones son iguales porque la cantidad de productos que es posible tener a la mano va de 0 a M y la cantidad de los que podemos ordenar tiene el mismo rango. Es decir,

$$X = \{0, 1, 2, \ldots M\} = A,$$

sin embargo, cuando el estado en que se encuentra el inventario es x, la máxima cantidad de productos que se puede ordenar es $M - x$ dada la capacidad limitada de almacenamiento. Esta observación la recogemos considerando subconjuntos de A formados por las acciones admisibles cuando el sistema se encuentra en cada estado x a los que denotaremos por $A(x)$

$$A(x) = \{0, 1, 2, \ldots M - x\}.$$

Si denotamos la demanda aleatoria al tiempo t por D_t, es claro que el estado del sistema al tiempo $t+1$ está dado por

$$x_{t+1} = (x_t + a_t - D_t)^+ = \max\{x_t + a_t - D_t, 0\},$$

esto se debe a que si la demanda excede a la existencia de productos a la mano $x_t + a_t$, entonces nuestro inventario queda vacío.

Si las D_t, $t = 1, 2, \ldots$, tienen distribución común en cualquier tiempo t dada por $p_j = \Pr[D_t = j]$, $j = 0, 1, \ldots$, entonces la probabilidad de transición es

$$p(y \mid x, a) = \begin{cases} 0 & si \quad y \geq x + a \\ p_{x+a-y} & si \quad x + a \geq y > 0 \\ \sum_{j=x+a}^{\infty} p_j & si \quad y = 0 \end{cases},$$

en donde se contempla que la probabilidad de pasar a un inventario igual a cero está dada por la probabilidad de que la demanda sea mayor o igual que los productos que

tenemos a la mano. Eliminamos el subíndice t debido a que esta probabilidad es igual en cualquier momento de transición.

Para darle forma a la función recompensa, tenemos que contemplar los distintos elementos que mencionamos anteriormente. Si el precio por artículo es de r pesos, entonces el ingreso esperado cuando se tiene un nivel de inventario u, esta dado por

$$g(u) = \sum_{j=0}^{u-1} rjp_j + ru \sum_{j=u}^{\infty} p_j.$$

El segundo sumando expresa el hecho de que si tenemos u productos y la demanda aleatoria es mayor o igual que esa cantidad, nuestro ingreso se reduce al valor de los u productos que pudimos vender.

Los gastos en que se incurre al ordenar la adquisición de una cantidad a de productos, incluyen un costo fijo de k pesos por orden y un costo variable $v(a)$ que depende de la cantidad de productos solicitados a la unidad de producción. Usualmente el costo variable está dado por $v(a) = ca$ donde c es la constante que representa el costo unitario de los productos que se manejan. Definimos entonces la función costo por ordenar a unidades como:

$$O(a) = \begin{cases} k + ca & si \quad a > 0 \\ 0 & si \quad a = 0 \end{cases}.$$

Se deben contemplar además los gastos por el almacenamiento y el manejo de los u productos que se tienen a la mano, que representaremos a través de la función $h(u)$. Por último, agregaremos una penalización de z unidades monetarias por cada producto demandado que no se haya podido surtir cuando se tienen u unidades a la mano, cuyo valor esperado está dado por

$$l(u) = \sum_{j=u+1}^{\infty} zjp_j.$$

De esta manera, si el estado observado es x y la acción elegida es a, el total de productos a la mano es $u = x + a$ y la función recompensa está dada por

$$r(x, a) = g(x + a) - O(a) - h(x + a) - l(x + a).$$

Como en el caso de las probabilidades de transición, las recompensas estarán definidas de la misma forma para cualquier tiempo t suponiendo que las variables económicas se mantienen constantes a lo largo del período de observación.

Una regla de decisión definida en términos de un límite inferior I que indica la mínima cantidad que es aceptable tener en nuestro almacén para surtir la demanda y un límite superior S que representa la máxima cantidad que es conveniente tener a la

mano, quedaría indicada por

$$d_t(x) = \begin{cases} S - x & si \quad x \leq I \\ 0 & si \quad x > I \end{cases},$$

para cualquier t. La determinación de los valores S e I se debe hacer de tal manera que la política $\pi = (d_1, d_2, \ldots d_{N-1})$ *maximice la recompensa total esperada.*

Ejemplo 1.2.2 ***Juego de tragamonedas.*** *Este ejemplo se trata de analizar cómo debe actuar un jugador que está apostando en una máquina donde coloca cierta cantidad de dinero cada vez que juega y existe una probabilidad, desconocida por él, de que la máquina le regrese una cantidad de dinero mucho mayor que la que pagó. El jugador puede tomar únicamente dos decisiones: jugar o no jugar.*

Conocer la probabilidad q de ganar, le daría al jugador elementos suficientes para decidir si es conveniente seguir apostando o no, de manera que uno de los problemas a resolver es cómo obtener una buena aproximación de esa probabilidad. Suponiendo que el jugador tiene ciertas creencias iniciales acerca del valor de q que se resumen en una función $f(q)$*, se puede utilizar estadística Bayesiana para* aprender *de lo ocurrido en cada juego e ir mejorando el conocimiento de la probabilidad de ganar. Así, si el apostador opta por jugar, el conjunto de estados será la colección de funciones de q que se pueden ir obteniendo con base en este proceso de aprendizaje. Por otro lado, si el apostador decide no jugar este proceso de aprendizaje no tendrá lugar, pero tampoco perderá ninguna cantidad de dinero.*

Podemos entonces considerar que tenemos dos cadenas de Markov (no controladas) $\left\{X_t^{(1)}\right\}$ *y* $\left\{X_t^{(2)}\right\}$ *con conjuntos de estados y probabilidades de transición propios cada una de ellas, que generan sucesiones de recompensas* $\left\{r^{(1)}\left(X_t^{(1)}\right)\right\}$ *y* $\left\{r^{(2)}\left(X_t^{(2)}\right)\right\}$. *La primer cadena corresponde a la decisión de no jugar y la segunda a la de jugar. En cada época de decisión t el apostador debe decidir a cuál de las dos cadenas le permite evolucionar, partiendo de que la otra se quedará en el estado en que se encuentra durante un período más. Una vez tomada esta decisión, la evolución de la cadena elegida no puede ser influenciada por él. Lo único que el jugador decide es, entonces, cuál cadena evolucionará en el período entre* t *y* $t+1$.

La cadena generada por la decisión de no jugar es trivial: tiene un único estado al que podemos llamarle 1, *y la transición es segura. es decir:*

$$X^{(1)} = \{1\}, \quad y \quad p^{(1)}(1 \mid 1) = 1.$$

Evidentemente, si el apostador no juega (es decir, si decide dejar evolucionar esta cadena) la recompensa sería $r^{(1)} \equiv 0$ *en cualquier época de decisión.*

Si decide jugar, supondremos que el apostador paga una cantidad c menor que 1, y tras jalar la palanca de la máquina, gana un peso con probabilidad q y cero pesos con probabilidad $1-q$. *El estado de la cadena en cada momento t resume lo que el apostador ha aprendido del valor de q a través de una función* $\dot{f} : [0,1] \to [0,1]$ *que se puede pensar como una asignación de* pesos *o ponderaciones para los distintos valores de q. Así, el espacio de estados es el conjunto de funciones de densidad que tienen soporte en el intervalo* $[0,1]$. *Entonces, escribiremos el espacio de estados de la segunda cadena como:*

$$X^{(2)} = \{f : f \text{ es una densidad con soporte en } [0,1]\}.$$

En un modelo Bayesiano, a la función f se le llama distribución a priori.

Puesto que la probabilidad de que el jugador gane una unidad es q, la recompensa esperada está dada por:

$$r^{(2)}(f) = \int_0^1 qf(q)dq - c = E_f[Q] - c,$$

donde Q denota la variable aleatoria con densidad f, y E_f *es el operador esperanza respecto a esa densidad.*

En este caso las transiciones tienen lugar entre densidades. En cada época de decisión contamos con una densidad a priori *f que hemos dado por buena aproximación durante el período que termina y queremos pasar una densidad modificada f ' (distribución* a posteriori*). Evidentemente, si en el juego anterior el apostador ganó, f ' pondrá mayor peso en valores grandes de q con relación al peso que les daba f. Análogamente, si perdió tendrá que darle más peso a valores pequeños de q. Podemos razonar informalmente el problema y escribir* $Q \approx q$ *en lugar de* $Q \in [q, q+dq]$, *para obtener:*

$$P(Q \approx q \mid \text{ganó}) = \frac{P(\text{gane} \mid Q \approx q)P(Q \approx q)}{\int_0^1 P(\text{gane} \mid Q \approx q)P(Q \approx q)dq} = \frac{qf(q)}{\int_0^1 qf(q)dq}. \tag{1.2.1}$$

De esta manera, la probabilidad de transición de la segunda cadena está dada por

$$p^{(2)}(f' \mid f) = \begin{cases} E_f[Q] & si \quad f' = \frac{qf(q)}{E_f[Q]} \\ 1 - E_f[Q] & si \quad f' = \frac{(1-q)f(q)}{1-E_f[Q]} \end{cases}.$$

En resumen, el proceso de decisión Markoviano que modela este problema tiene los siguientes elementos:

El conjunto de épocas de decisión es $T = \{1, 2, \ldots, N\}$ *con* $N \leq \infty$.

El espacio de estados está formado por parejas:

$$X = X^{(1)} \times X^{(2)} = \{(1, f) : f \text{ es una densidad con soporte en } [0,1]\}.$$

El espacio de acciones viables es igual para todo estado x y está dado por : $A = A(x) = \{c_1, c_2\}$ donde c_i corresponde a dejar que evolucione la cadena i, $i = 1, 2$.

La probabilidad de transición está dada por:

$$p\left(\left(1, f'\right) \mid (1, f), a\right) = \begin{cases} p^{(1)}(1 \mid 1) & si \quad a = c_1 \\ p^{(2)}(f' \mid f) & si \quad a = c_2 \end{cases}.$$

Finalmente, la función recompensa se define como:

$$r_t((1, f), a) = \begin{cases} r^{(1)}(1) & si \quad a = c_1 \\ r^{(2)}(f) & si \quad a = c_2 \end{cases}.$$

Sin embargo, considerar el espacio de estados como el conjunto de todas las densidades con soporte en $[0, 1]$, hace el problema computacionalmente imposible de resolver. Hay dos posibles caminos para hacerle frente a esta limitación:

1. **Ejemplo 1.2.3** *Escoger una familia paramétrica de densidades que sea cerrada bajo el cálculo en 1.2.1. Por ejemplo si el estado inicial $x_1^{(2)} = f_1$ es una densidad beta, entonces $x_n^{(2)}$ será una densidad beta para toda n y puede describirse completamente por el número de pruebas, el número de juegos ganados y los parámetros iniciales de la densidad. Supongamos que los parámetros iniciales son α y β, es decir:*

$$f_1(q) = \frac{\Gamma(\alpha + \beta)}{\Gamma(a)\Gamma(\beta)} q^{\alpha - 1}(1 - q)^{\beta - 1}.$$

Si después de n juegos se han tenido k triunfos, entonces el estado $x_{n+1}^{(2)}$ será una densidad beta con parámetros $\alpha + k$ y $\beta + n - k$. Entonces, lo único que hay que determinar son los parámetros α y β . El espacio de estados se puede representar como $X^{(2)} = (0, \infty) \times (0, \infty)$ correspondiente a los parámetros de la distribución beta o como $X^{(2)} = \{0, 1, ...\} \times \{0, 1, ...\}$ correspondiente al número de jugadas ganadas y el de jugadas perdidas.

La otra vía es representar al espacio de estados como $X^{(2)} = \{0, 1, ...\} \times \{0, 1, ...\}$ correspondiente al número de jugadas ganadas y pérdidas, y usar la igualdad

$$f(q) = \frac{q^k (1 - q)^{n-k} f_1(q)}{\int_0^1 q^k (1 - q)^{n-k} f_1(q) dq}$$

para calcular la f en el estado $(k, n - k)$ antes de cada jugada.

1.3 Política óptima de reclamos en un seguro multiperíodo

Entrando ya a los problemas de seguros que nos interesa analizar en este trabajo, expondremos aquí las características generales del primero de ellos.

Supongamos que una persona contrata un seguro de automóvil que abarca N períodos. Al inicio de cada período la persona paga una cuota correspondiente a la prima temporal. El asegurado sabe que si realiza reclamos, la aseguradora lo clasificará en un nivel de mayor riesgo en el siguiente período, lo que repercutirá en una prima más alta. En cambio, si no incurre en reclamos, para el siguiente período será ubicado en un nivel de prima más bajo. De esta manera, al sufrir un accidente, el asegurado debe tomar en consideración este hecho para decidir si hace la reclamación correspondiente o no. Es decir, a diferencia de los contratos que abarcan un sólo período, en este caso el comprador del seguro debe encontrar un nivel de daños a partir del cual le conviene reclamar, considerando que si hace todos los reclamos, a largo plazo su utilidad será menor que si selecciona cuáles cobrar.

Este enfoque puede extenderse a otras áreas de seguro de daños que tienen una estructura similar, por ejemplo ciertos tipos de seguros médicos y de seguros contra robos.

Inicialmente, supondremos que el contrato es fijo y que el asegurado considera que dicho contrato es su mejor opción. Las categorías de riesgo en el contrato serán representadas por el conjunto $\{1, 2, \ldots, J\}$, donde 1 representa el menor riesgo y J el máximo. Si un asegurado que se encuentra en la categoría j hace un reclamo durante el período vigente, entonces será ubicado en la categoría $j+1$ al siguiente período. Por el contrario, si no realiza ningún reclamo, el siguiente período será colocado en la categoría $j-1$. Cuando un asegurado que se encuentra en el máximo nivel de riesgo hace un reclamo, en el siguiente período será ubicado en un nivel ficticio denotado por $J+1$, que representa que ha quedado fuera del seguro.

La variable aleatoria $X_{j,t}$ indica el monto del daño ocurrido durante el período t a un asegurado que se encuentra en la categoría j. Si suponemos que la distribución de los daños es independiente del período en el que ocurren, podemos eliminar el subíndice t y escribir solamente X_j.Cada una de las variables aleatorias X_j toma valores en el intervalo $[0, M]$ donde M es el monto máximo de daños que cubre el seguro contratado. $X_j = 0$ representa el hecho de que no ocurrió un siniestro durante el período. Denotaremos la distribución de estas variables aleatorias por $F_j(x) = P[X_j \leq x]$. Como el subíndice j representa riesgo, supondremos que

$$F_{j+1}(x) \geq F_j(x) \quad \text{para toda } x.$$

Si F_j es una distribución absolutamente continua para cada j, denotaremos por f_j a la densidad correspondiente, es decir:

$$F_j(x) = \int_{-\infty}^{x} f_j(y)\, dy.$$

De lo anteriormente expuesto se deduce que las decisiones que puede tomar el asegurado en caso de que ocurra un siniestro, son exclusivamente dos: reclamar o no reclamar. Para tomar la decisión, el asegurado observa el monto del daño x y el nivel de riesgo j en que se encuentra ubicado, es decir, los estados serán parejas de la forma (x, j).

Si en el período t ocurre un reclamo cuyo monto es de x pesos, la aseguradora paga una indemnización dada por $I_{jt}(x)$ cuando el asegurado se encuentre en la categoría j. Con el fin de no representar un incentivo para incurrir en pérdidas, la función de indemnización satisface las condiciones

$$0 \leq I_{jt}(x) \leq x, \quad \forall\, j, t, \text{ y } x \neq 0,$$

y para $x = 0$,

$$I_{jt}(0) = 0.$$

El precio del contrato que paga una indemnización I_{jt}, es decir, el monto de la prima, será denotado por r_j. Supondremos que

$$r_j > r_{j-1},$$

para cualquier tiempo t, en virtud de que a mayor j mayor riesgo.

Sea C_t el ingreso del asegurado durante el período t. La recompensa que recibirá un asegurado ubicado en el nivel j que decida no reclamar un daño cuyo monto es x, será entonces $C_t - r_j - x$, y si decide hacer el reclamo correspondiente será $C_t - r_j - x + I_{jt}(x)$.

En este problema buscaremos maximizar una función de las recompensas esperadas que contemple la utilidad que al asegurado le reporta la contratación del seguro. Para ello, tenemos que considerar una función utilidad $u(\cdot)$ que refleje los beneficios de la cobertura tanto en el terreno económico como en el aspecto de la seguridad que le brinda al asegurado contar con el seguro. La discusión acerca de la relación entre utilidad y seguros la daremos en el siguiente capítulo.

En resumen, los elementos del modelo para el problema de buscar una política de reclamos óptima, son:

- Conjunto de épocas de decisión:

$$T = \{1, 2, \ldots, N\}.$$

- Espacio de estados:

$$X = \{(x, j) \mid x \in [0, M],\ j \in \{1, 2, \ldots J + 1\}\}.$$

- Espacio de acciones:

$$A_{(x,j)} = \{0,1\} \quad \text{para } x \neq 0,$$

donde 0 representa la acción de no reclamar y 1 la de hacer el reclamo correspondiente; y

$$A_{(0,j)} = \{0\}.$$

- Probabilidades de transición:

$$p_t\left[(B,i) \mid (x,j),0\right] = \begin{cases} \int_{[0,M]\cap B} f_{j-1}(x)\,dx & \text{si } i = j-1 \\ 0 & \text{si } i \neq j-1 \end{cases}$$

y

$$p_t\left[(B,i) \mid (x,j),1\right] = \begin{cases} \int_{[0,M]\cap B} f_{j+1}(x)\,dx & \text{si } i = j+1 \\ 0 & \text{si } i \neq j+1 \end{cases},$$

para toda t.

- Función de recompensa:

$$r_t\left((x,j),a\right) = \begin{cases} u(C_t - r_j - x) & \text{si } a = 0 \\ u(C_t - r_j - x + I_{jt}(x)) & \text{si } a = 1 \end{cases}.$$

En el capítulo 4 resolvemos el problema de determinar la política óptima de reclamos y nos planteamos un problema adicional: ¿cuál es el tipo de contrato óptimo para un asegurado que lleva a cabo una política como la que resulta óptima?

1.4 Determinación de la prima óptima cuando se desconoce la distribución de los reclamos

El segundo problema que será analizado es el problema de la determinación de la prima óptima por parte de una aseguradora. Para determinar el monto de las primas que deberá cobrar por un cierto tipo de póliza de seguro, una compañía aseguradora toma en cuenta la distribución de probabilidad de los reclamos que ocurrirán en el período que inicia. Si se supone conocida la distribución de los reclamos y se mantiene fija a lo largo de todos los períodos en observación, entonces la aseguradora no está utilizando la información sobre cómo ha sido el proceso de reclamos para aprender algo sobre los reclamos futuros. Una mejor alternativa es ir construyendo, al principio de cada período, una distribución de probabilidad de los reclamos con base en la información que brinda lo ocurrido en el período anterior y determinar la prima que se cobrará con base en esa distribución. En el capítulo 5 de este trabajo, expondremos brevemente

los conceptos básicos de estadística bayesiana que serán necesarios para resolver el problema de la determinación de primas en cada período de acuerdo a esta óptica.

La aseguradora obtiene información tanto del número de reclamos que se presentaron en el período anterior, como del monto (en pesos) de dichos reclamos. Vamos a considerar que la distribución de los reclamos y de su monto es desconocida en el sentido de que se conoce la forma paramétrica de dicha distribución, pero no se conoce el valor concreto del parámetro que la determina. Para encontrar el valor de dicho parámetro, partimos de una conjetura inicial y utilizamos la información sobre lo ocurrido en el período anterior para modificar esta conjetura en cada época de decisión (como se hizo en el ejemplo del juego de tragamonedas).

Se considera también que la demanda del seguro en cada período está determinada por el monto de la prima, misma que se fija con base en la distribución de dos variables aleatorias: X que indica el número de reclamos y Y que representa el monto de los mismos.

A su vez, como el monto de la prima fijada al tiempo t determina el número de seguros que se venderán en el siguiente período, también afecta la distribución de los reclamos en ese período. En este sentido, la prima en el período $t+1$ depende de la información que brinde el proceso de reclamos ocurridos con anterioridad. Es decir, los distintos montos de las primas en cada período, están interrelacionados.

Sea N el número de pólizas vendidas del tipo de seguro que se está analizando. Supondremos que los períodos son suficientemente pequeños para que a lo más ocurra un reclamo de cada asegurado durante ese lapso. Sea X_i la variable aleatoria dada por

$$X_i = \begin{cases} 1 & \text{el } i-\text{ésimo asegurado hace un reclamo} \\ 0 & \text{en caso contrario} \end{cases}.$$

Si conociéramos la probabilidad

$$p = \Pr\left[\text{ocurra un reclamo}\right],$$

entonces la distribución de $X = \sum_{1}^{N} X_i$ sería una binomial con parámetros N y p, dado que cada X_i se distribuye como Bernoulli con parámetro p. Queremos entonces desarrollar un proceso de aprendizaje sobre p.

Por otro lado, sea Y_i el monto del $i-$ésimo reclamo. Supondremos que las variables aleatorias Y_i son independientes y que se distribuyen como exponencial con parámetro w. De esta manera, para un número x de reclamos, el monto total de los reclamos Y se distribuirá como una función gama con parámetros x y w, siendo w el parámetro desconocido w. Desarrollaremos también un proceso de aprendizaje sobre w,que representa el tiempo promedio que transcurre antes de que ocurra un reclamo

Los dos procesos de aprendizaje se desarrollan a través de funciones f_p y f_w que asignan pesos diferentes a los distintos valores de los parámetros desconocidos de acuerdo a la información que brinde el proceso de reclamos ocurrido en el período anterior. La conjetura inicial de la aseguradora sobre el parámetro p se resume en una distribución beta f_p con parámetros r y n, donde r representa la creencia inicial sobre el número de reclamos y n la creencia inicial acerca del número de pólizas que se venderán.

Por otro lado, la conjetura inicial de la compañía aseguradora sobre el parámetro w se resume en que f_w es una distribución gama con parámetros α y β, donde α es la creencia inicial sobre el número de reclamos y β sobre el monto de los mismos.

Considerando estos procesos de aprendizaje, los estados formados por parejas (x, y) se transforman vectores (α, β, r, n) donde se recogen los parámetros mencionados en el párrafo anterior.

Si en el siguiente período ocurrieran x reclamos en N pólizas, entonces el valor de p se determinará de acuerdo a una distribución beta $\mathbf{B}(p \mid r + x, n + N)$. Análogamente, el valor posterior de w está dado por una distribución gama $\mathbf{g}(w \mid \alpha + x, \beta + y)$ suponiendo que y fue el monto de los x reclamos ocurridos. Con base en esa información, la aseguradora decidirá el monto π de la prima para el período que inicia y se tendrá un nuevo número $N(\pi)$ de pólizas vendidas para el análisis siguiente.

Entonces, los elementos del modelo son:

- Conjunto de épocas de decisión:

$$T = \{1, 2, \ldots, K\},$$

  para alguna K finita.

- Espacio de estados:

$$X = \{(\alpha, \beta, r, n) \mid \alpha, r, n \in \mathbb{N}, \beta \in \mathbb{R}\}.$$

- Espacio de acciones:

$$A = [0, \infty),$$

  que es, en principio, el rango del monto π de la nueva prima.

- Probabilidades de transición: Si se venden $N(\pi)$ pólizas y ocurren x reclamos cuyo monto suma y, el nuevo estado será $(\alpha + x, \beta + y, r + x, n + N(\pi))$. Analizaremos la probabilidad de que esto ocurra en el capítulo 5 de este trabajo.

- Función de recompensa: construiremos una función $\Psi(\pi \mid \alpha, \beta, r, n)$ que represente la utilidad de la aseguradora en el período dada por:

$$r\left(\left(\alpha,\beta,r,n\right),\pi\right)=\Psi\left(\pi\mid\alpha,\beta,r,n\right)=\pi N\left(\pi\right)-E\left[Y\mid\alpha,\beta,r,n\right].$$

Capítulo 2

Teoría de utilidad y seguros

2.1 Introducción

Un sistema de seguros es un mecanismo para reducir los impactos financieros adversos producidos por eventos aleatorios, ofreciendo la cobertura de montos económicos razonables en caso de que ocurra la pérdida de un bien o el fallecimiento de alguna persona. El sustento elemental de un sistema de seguros radica en cobrar ciertas cantidades de dinero relativamente pequeñas (primas) a un grupo amplio de personas durante un período de tiempo, de tal manera que la suma de esas cantidades permita cubrir las reclamaciones que se presenten durante ese período por siniestros ocurridas. Por esta razón, el monto de las primas puede determinarse a través de una función de la esperanza del número de pérdidas, pero también es posible calcularlas en función de la cantidad que los asegurados están dispuestos a pagar por el bien que desean proteger.

La teoría de utilidad es un instrumento para la toma de decisiones en presencia de incertidumbre, que en particular tiene aplicaciones en la determinación de primas de seguros para distintos tipos de pólizas.

En este capítulo expondremos brevemente los conceptos básicos de esta teoría para referirnos a dichos conceptos en los siguientes capítulos.

2.2 Teoría de la utilidad

El valor de un proyecto económico cuyo resultado es aleatorio, puede identificarse con el valor esperado de su resultado económico. Siguiendo esta lógica, para la toma de decisiones la distribución de probabilidad de los resultados puede reemplazarse por un número único dado por la esperanza de la variable aleatoria del resultado económico del proyecto. De acuerdo a este principio, a una persona le daría lo mismo asumir

una pérdida aleatoria de X pesos que pagar una cantidad $E[X]$ para ser liberado de la posible pérdida. De manera similar, una persona estaría dispuesta a pagar una cantidad $E[Y]$ para participar en un juego en el que la ganancia aleatoria es de Y unidades monetarias.

Sin embargo, para muchas personas el monto de la riqueza en riesgo y otros aspectos de la distribución de probabilidad de los resultados, tienen influencia en sus decisiones, como se ve en el siguiente ejemplo:

Ejemplo 2.2.1 *Supongamos que la probabilidad de que ocurra una pérdida producida por un accidente es fija e igual a 0.1, mientras que la probabilidad de que tal pérdida no ocurra es de 0.9. Consideremos distintos montos para la posible pérdida como se muestra en el siguiente cuadro:*

Caso	*Pérdidas posibles*	*Pérdida esperada*
1	1	.1
2	1000	100
3	1 000 000	100 000

En el primer caso, la pérdida de 1 peso no es significativa, de manera que la persona estaría más dispuesta a asumir la pérdida que pagar 10 centavos por un seguro. Sin embargo, en el caso tres la pérdida es tan grande que es de esperarse que la persona opte por pagar el equivalente a su pérdida esperada para ser protegido por un seguro. Así, en dos proyectos económicos con la misma distribución de probabilidad, la actitud del asegurado potencial es radicalmente distinta.

Veamos entonces otra forma de calcular el valor que tiene un proyecto para una persona. Si denotamos la riqueza involucrada en un proyecto económico por w, diremos que el valor que le da el tomador de decisiones a esa riqueza es una función $u(w)$ a la que llamaremos función de utilidad. Así, si las preferencias de la persona satisfacen ciertos requerimientos de consistencia, la función de utilidad $u(w)$ brinda un criterio para la toma de decisiones dado por: preferirá la distribución de X a la de Y si $E[u(X)] > E[u(Y)]$. y será indiferente si $E[u(X)] = E[u(Y)]$.

Al usar la teoría de utilidad para tomar decisiones, es conveniente tener en cuenta las siguientes observaciones:

1. La teoría de utilidad es elaborada bajo la suposición de que existen preferencias por algún evento, así como por su probabilidad de ocurrencia.

2. Las preferencias se preservan cuando se aplica a la utilidad una transformación lineal creciente (creciente, para que la desigualdad se preserve). La demostración

de esta afirmación se basa fundamentalmente en la linealidad de la esperanza. Por ejemplo, si w representa la riqueza y

$$u^*(w) = au(w) + b \qquad a, b \in \mathbb{R} \quad \text{y} \quad a > 0,$$

entonces,

$$E[u(X)] > E[u(Y)]$$

es equivalente a

$$E[u^*(X)] > E[u^*(Y)].$$

3. Si la transformación lineal de la función de utilidad es como se describió anteriormente y si $E[X] = \mu_X$ y $E[Y] = \mu_Y$, tenemos que

$$E[u^*(X)] = a\mu_X + b > E[u^*(Y)] = a\mu_Y + b$$

se cumple, si y solo si

$$\mu_X > \mu_Y.$$

Esto quiere decir que para transformaciones lineales y crecientes, las preferencias por las distribuciones de los resultados se preservan al comparar los valores esperados de las distribuciones.

2.3 Seguro y utilidad

En seguros, el monto del pago de la prima se determina adoptando un principio económico de decisión aceptable tanto por el asegurado como por el asegurador. Esto es posible para una póliza de seguros en la cual la prima establecida por el asegurador es menor que el monto máximo que esta dispuesto a pagar el propietario del bien a asegurar, monto que será determinado más adelante.

Por ahora, consideraremos que el tomador de decisiones es la persona que va a asegurar una propiedad expuesta a un daño aleatorio, o a la pérdida total, durante el siguiente periodo. El monto de la pérdida es una variable aleatoria denotada por X y supondremos que su distribución es conocida.

Al valor esperado $E(X) = \mu$ le llamaremos la *prima neta o pura* correspondiente a la póliza en consideración durante un período. En la práctica, el asegurador agrega un valor de recargo a la prima neta, debido a impuestos, gastos de administración, etcétera. A esta nueva prima H la llamaremos *prima recargada* y podemos obtenerla mediante una función de la prima neta de la siguiente manera:

$$H = \mu(1 + \theta) + c \qquad \theta > 0 \ \text{y} \ c > 0.$$

El factor $\mu\theta$ corresponde a los gastos en que incurre el asegurador por las variaciones que se presentan respecto a las pérdidas esperadas, es decir, debido a que la ocurrencia de reclamos se desvíe de la esperada. La constante c corresponde a gastos fijos que no varían según el monto de las pérdidas.

Ahora aplicaremos la teoría de utilidad para el tomador de decisiones. Desde el punto de vista del asegurado, el valor extremo de una prima G estaría determinado por una ecuación donde la utilidad por asegurar un bien (lado izquierdo de la siguiente ecuación) sea igual al valor esperado de la pérdida en caso de asumir el riesgo sin protección (lado derecho):

$$u(w - G) = E[u(w - X)], \tag{2.3.1}$$

donde G es la prima que se pagaría por asegurar el bien.

Si la función de utilidad es lineal del tipo $u(w) = bw + d$, la ecuación anterior toma la forma

$$\begin{aligned} u(w - G) &= b(w - G) + d \\ &= E[u(w - X)] \\ &= E[b(w - X) + d], \end{aligned}$$

es decir,

$$\begin{aligned} b(w - G) &= b(w - \mu) \\ G &= \mu. \end{aligned}$$

Cuando se de la igualdad el asegurado será indiferente a las opciones de pagar por el seguro o asumir él mismo el riesgo. Recordemos que en ausencia de un subsidio, el asegurador debe cobrar una cantidad por encima de sus pérdidas esperadas, es decir un monto mayor a $E[X]$, por lo que en caso de que se cumpla la ecuación (2.3.1) para una utilidad lineal, el contrato de seguro es inviable.

En un modelo más realista, podemos suponer que la función de utilidad es una función en que a mayor riqueza corresponde mayor utilidad, Además, se ha observado que para muchos tomadores de decisiones, cada incremento del mismo tamaño en la riqueza, corresponde a un menor incremento en su utilidad mientras mayor sea la riqueza. Tal función de utilidad en economía, se conoce como la *utilidad marginal.* Gráficamente este tipo de función de utilidad es del tipo cóncava creciente (ver figura 1), cuya propiedad es que la primera derivada de $u(w)$ es positiva y la segunda es negativa.

En resumen, las consideraciones anteriores nos llevan a utilizar como función de utilidad una función $u : \mathbb{R} \to \mathbb{R}$ continua con las siguientes propiedades:

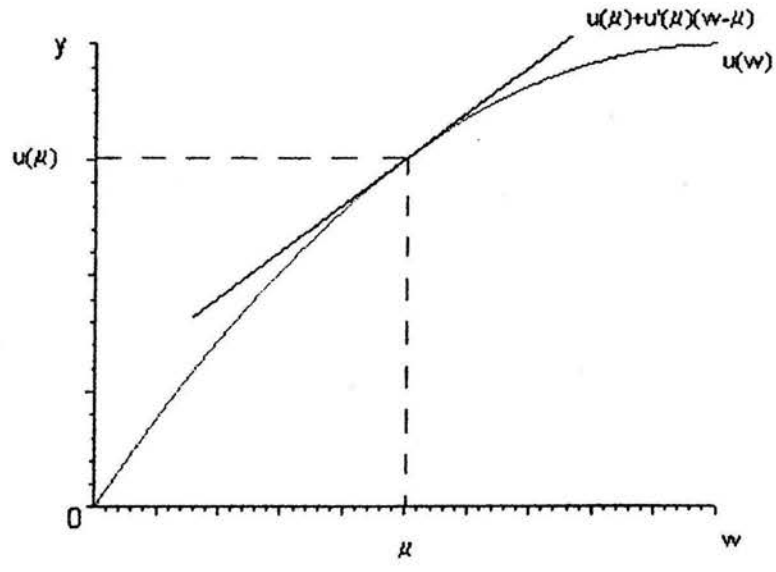

Figure 2.1: Comportamiento gráfico de la desigualdad de Jensen

1. $u'(\cdot) > 0$
2. $u''(\cdot) < 0$

Para este tipo de funciones podemos usar la desigualdad de Jensen dada en el siguiente Teorema:

Teorema 2.3.1 *Desigualdad de Jensen. Sea $u(w)$ una función creciente de $\mathbb{R}$ en $\mathbb{R}$ tal que si $u''(w) < 0$ y X es una variable aleatoria, entonces $E[u(X)] \leq u[E(X)]$.*

Gráficamente podemos interpretar el Teorema anterior de la siguiente forma: en el punto de la gráfica donde $w = E(X) = \mu$, tomamos la mejor aproximación lineal de $u(\mu)$ dada por

$$y = u(\mu) + u'(\mu)(w - \mu),$$

y por la concavidad de la función, la gráfica queda debajo de la línea tangente y, es decir, $u(w) \leq y$ para toda w. Por último si w es reemplazada por X y tomamos la esperanza de ambos lados de la desigualdad,

$$E[u(X)] \leq u[E(X)].$$

Si aplicamos la desigualdad de Jensen a la ecuación (2.3.1), como lo hicimos en el caso lineal, tenemos que

$$u(w - G) = E[u(w - X)] \leq u(w - \mu).$$

Debido a que $u(w) > 0$ y que $u(w)$ es una función creciente, tenemos que

$$w - G \leq w - \mu,$$

de donde se desprende finalmente que

$$G \geq \mu.$$

La igualdad se da solo si la variable aleatoria X es constante.

En términos económicos, lo obtenido anteriormente significa que si $u(w)$ es una función cóncava y creciente, el tomador de decisiones estará dispuesto a asegurarse pagando un monto mayor a la pérdida esperada. A un tomador de decisiones con esta disposición se le denominará *adverso al riesgo.* Si G es al menos igual a la prima establecida por el asegurado hay una oportunidad clara de establecer un contrato de seguro.

Desde el punto de vista del asegurador la ecuación para definir el mínimo aceptable H que estaría dispuesto a cobrar por asumir el riesgo es:

$$u_1(w_1) = E[u_1(w_1 + H - X)]. \tag{2.3.2}$$

El lado izquierdo corresponde a la utilidad que le otorga disponer de su riqueza w_1.El lado derecho es el valor esperado de la utilidad que obtiene de cobrar la prima H y pagar la cantidad aleatoria X. Cuando la igualdad se cumple, la ecuación nos dice que el asegurador es indiferente entre su posición actual o proveer el seguro de X por una prima de H. Usando la desigualdad de Jensen obtenemos:

$$u_1(w_1) = E[u_1(w_1 + H - X)] \leq u_1(w_1 + H - \mu),$$

y con el mismo razonamiento que para el asegurado, $H \geq \mu$. Si la G del asegurado es tal que $G \geq H \geq \mu$, es posible una póliza de seguros aceptable para ambas partes.

Para efectos prácticos, vamos a ver aquí las ventajas de trabajar con una función de utilidad exponencial

$$u(w) = -e^{-\alpha w} \qquad \alpha > 0.$$

Primero vemos que

$$u'(w) = \alpha e^{-\alpha w} > 0,$$

y

$$u''(w) = -\alpha^2 e^{-\alpha w} < 0.$$

En segundo lugar,

$$E[u(X)] = E[-e^{-\alpha X}] = -M_X(-\alpha),$$

que es la función generatriz de momentos con $t = -\alpha$ multiplicada por -1.

En tercer lugar, las primas no dependen en la riqueza del tomador de decisiones. Esto se verifica sustituyendo la función de utilidad exponencial en (2.3.1). Esto es,

$$\begin{aligned} -e^{-\alpha(w-G)} &= E[-e^{-\alpha(w-X)}] \\ e^{\alpha G} &= M_X(\alpha) \\ G &= \frac{\ln M_X(\alpha)}{\alpha}, \end{aligned}$$

que no depende de w.

La verificación para el asegurador en este caso particular, se hace en la ecuación (2.3.2) con α_1

$$\begin{aligned} -e^{-\alpha_1 w_1} &= E[-e^{-\alpha_1(w_1+H-X)}] \\ -e^{\alpha_1 w_1} &= -e^{-\alpha_1(w_1+H)} M_X(\alpha_1) \\ H &= \frac{\ln M_X(\alpha_1)}{\alpha_1} \end{aligned}$$

Ejemplo 2.3.2 *Supóngase que el tomador de decisiones usa una función de utilidad dada por* $u(w) = -e^{-5w}$. *Además, el tomador de decisiones tiene dos proyectos económicos aleatorios disponibles. El resultado del primero, denotado por* X, *tiene distribución* $N(5,2)$, *y el segundo, denotado por* Y, *se distribuye* $N(6,2.5)$. *¿Que proyecto preferirá?*

Solución 2.3.3 *Los valores esperados de la utilidad de cada uno de los proyectos, son:*

$$E[u(X)] = E\left[-e^{-5X}\right] = -M_X(-5) = -e^{\left[-5(5)+\frac{5^2(2)}{2}\right]} = -1$$

y

$$E[u(Y)] = E\left[-e^{-5Y}\right] = -M_Y(-5) = -e^{\left[-5(6)+\frac{5^2(2)}{2}\right]} = -e^{1.25}.$$

Por lo que

$$E[u(X)] = -1 > E[u(Y)] = -e^{1.25},$$

y se prefiere la distribución de X *sobre la de* Y.

Es importante notar aquí que, aunque $E[X] = 5 < 6 = E[Y]$ se prefirió la distribución de Y debido al efecto de la función de utilidad específica. Al usar el principio del valor esperado para discriminar entre las distribuciones, estamos presuponiendo que el tomador de decisiones es adverso al riesgo.

Otra familia de funciones de utilidad, es la de potencias fraccionarias, dada por:

$$u(w) = w^{\gamma}; \qquad w > 0 \text{ y } 0 < \gamma < 1.$$

que cumple:

$$u'(w) = \gamma w^{\gamma-1} > 0$$

y

$$u''(w) = \gamma(\gamma - 1) w^{\gamma-2} < 0.$$

En esta familia, las primas dependen de la riqueza del tomador de decisiones de una manera que puede ser muy realista en muchos casos.

La familia de las funciones de utilidad cuadráticas es:

$$u(w) = w - \alpha w^2, \qquad w < (2\alpha)^{-1}, \alpha > 0,$$

con

$$u'(w) = 1 - 2\alpha w > 0 \quad \text{cuando} \quad w < (2\alpha)^{-1},$$

y

$$u''(w) = -2\alpha < 0.$$

Este tipo de funciones presenta el problema de que la prima depende de la riqueza del asegurado y, a diferencia de las anteriores, en la mayoría de los casos es poco realista. El problema es que con la misma distribución, a una mayor riqueza le corresponde un mayor cargo en la prima, contraviniendo el hecho de que si esto sucede es mas viable que el tomador de decisiones asuma el riesgo él mismo. Consecuentemente este tipo de funciones no serán consideradas por tomadores de decisiones para los que asumir el riesgo por cuenta propia sea menos preferible mientras mayor sea su riqueza.

Ejemplo 2.3.4 *La probabilidad de que una propiedad no sea dañada en el siguiente periodo es de 0.75. La función de densidad de la pérdida esta dada por:*

$$f(x) = 0.25\left[0.01e^{-0.01x}\right] \qquad x > 0.$$

El dueño de la propiedad tiene una función de utilidad dada por:

$$u(w) = -e^{-0.005w}.$$

Calcular la pérdida esperada y la prima máxima que el propietario pagará.

Solución 2.3.5 *La pérdida esperada esta dada por*

$$E[X] = 0.75(0) + 0.25\int_0^{\infty} x\left(0.01e^{-0.01x}\right) dx = 25,$$

y aplicando la ecuación para el asegurado (2.3.1) obtenemos:

$$u(w-G) = 0.75u(w) + \int_0^{\infty} u(w-x) f(x)\, dx,$$

de donde

$$G = 44.63;$$

por lo que el tomador de decisiones pagará hasta 44.63-25=19.63 por encima de su pérdida esperada para comprar el seguro.

2.4 Seguro óptimo para un período (Prueba de Arrow)

Para terminar este capítulo, veremos una prueba para determinar qué tipo de póliza le conviene comprar a un asegurado para maximizar su utilidad esperada, en el caso en que el seguro aplica sólo por un período.

El tomador de decisiones se enfrenta al problema de comprar un seguro en el que la cantidad que recibirá por un reclamo correspondiente a un daño x, es de $I(x)$ pesos, con $0 \leq I(x) \leq x$. La desigualdad anterior se impone para evitar que el asegurado sea motivado a incurrir en la pérdida y cobrar el seguro. Suponemos que el asegurado puede comprar cualquier tipo de póliza en la que el monto que reciba por un reclamo no exceda al monto esperado del daño, es decir, que $E[I(X)] \leq E[X]$. Supondremos además que el tomador de decisiones ha fijado el monto, denotado por P, a pagar por asegurarse.

La pregunta es, ¿cuál póliza de la clases de pólizas viables deberá ser comprada para maximizar la utilidad esperada del tomador de decisiones en un período?

Una subclase de la clase de pólizas viables es la de deducible d fijo, definida como:

$$I_d(x) = \begin{cases} 0 & x < d \\ x-d & x \geq d, \end{cases} . \tag{2.4.3}$$

Este tipo de pólizas a menudo se denominan seguro de *exceso de pérdida* o *stop loss*, porque el pago del reclamo no se hace hasta que la pérdida exceda el monto del deducible d.

En esta sección vamos a considerar que la prima P es igual a los reclamos esperados, por lo que:

$$P = \int_d^{\infty} (x-d) f(x)\, dx$$

Lema 2.4.1 *Si* $u''(w) < 0$, *entonces*

$$u(w) - u(z) \leq (w - z)\, u'(z).$$

Demostración. *Como se ve en la figura 1, la línea tangente a* $u(w)$ *en el punto* $[w_0, u(w_0)]$ *tiene la ecuación*

$$y - u(w_0) = u'(w_0)(w - w_0),$$

y está por encima de la función $u(w)$, *excepto en el punto de tangencia. De allí se tiene la desigualdad*

$$u(w) - u(w_0) \leq (w - w_0)\, u'(w_0).$$

El mismo argumento puede repetirse para cualquier valor z. ■

Teorema 2.4.2 *Si un tomador de decisiones*

- *tiene una riqueza* w;
- *es adverso al riesgo, es decir, la función de utilidad de la riqueza es tal que* $u'(w) > 0$ *y* $u''(w) < 0$;
- *enfrenta una pérdida aleatoria* X;
- *gastará un monto* P *en un seguro, donde* $0 < P \leq E[X] = \mu$.

Y si el asegurador

- *ofrece las pólizas viables cuyo monto por un reclamo* x *es* $I(x)$, *con* $0 \leq I(x) \leq x$, *y*
- *ofrece la compra de una póliza por el monto de la pérdida esperada,* $E[I(X)]$;

entonces la utilidad esperada del tomador de decisiones será máxima comprando una póliza de seguros que, por un reclamo x, *pague la cantidad*

$$I_{d^*}(x) = \begin{cases} 0 & x < d^* \\ x - d^* & x \geq d^* \end{cases},$$

donde d^* *es la solución de*

$$P - \int_d^\infty (x - d)\, f(x)\, dx = 0.$$

Demostración. Por el lema anterior, se tiene que

$$u(w - x + I(x) - P) - u(w - x + I_{d^*}(x) - P) \leq [I(x) - I_{d^*}(x)]\, u'(w - d^* - P). \tag{2.4.4}$$

Además reclamamos si

$$[I(x) - I_{d^*}(x)]\, u'(w - x + I_{d^*}(x) - P) \leq [I(x) - I_{d^*}(x)]\, u'(w - d^* - P). \tag{2.4.5}$$

Para establecer la última desigualdad, se deben contemplar tres casos:

Caso I. $I_{d^*}(x) = I(x)$.

Aquí la igualdad se satisface con 0 de ambos lados.

Caso II. $I_{d^*}(x) > I(x)$.

En este caso $I_{d^*}(x) > 0$ y de (2.4.3), $d^* = x - I_{d^*}(x)$ .Por lo que la igualdad se satisface con los dos valores iguales a $[I(x) - I_{d^*}(x)]\, u'(w - d^* - P)$.

Caso III. $I_{d^*}(x) < I(x)$.

En este caso $I(x) - I_{d^*}(x) > 0$.De (2.4.3) obtenemos $I_{d^*}(x) - x \geq -d^*$, e $I_{d^*}(x) - x - P \geq -d^* - P$.

Por lo cual,

$$u'(w - x + I_{d^*}(x) - P) \leq u'(w - d^* - P).$$

Debido a que la segunda derivada de $u(x)$es negativa y $u'(x)$es una función decreciente.

Por lo tanto, en cada caso,

$$[I(x) - I_{d^*}(x)]\, u'(w - x + I_{d^*}(x) - P) \leq [I(x) - I_{d^*}(x)]\, u'(w - d^* - P),$$

estableciendo la desigualdad 2.4.5.

Ahora, combinando las desigualdades 2.4.4 y 2.4.5, y tomando esperanza, tenemos

$$\begin{aligned} & E[u(w - X + I(X) - P)] - E[u(w - X + I_{d^*}(X) - P)] \\ \leq\ & E[I(X) - I_{d^*}(X)]\, u'(w - d^* - P) = (P - P)\, u'(w - d^* - P) = 0, \end{aligned}$$

por lo que,

$$E[u(w - X + I(X) - P)] \leq E[u(w - X + I_{d^*}(X) - P)],$$

es decir, la utilidad esperada será máxima seleccionando $I_{d^*}(x)$, la póliza stop-loss. ■

En el siguiente ejemplo se hace uso del deducible para determinar la prima máxima a pagar, usando el principio del valor esperado.

Ejemplo 2.4.3 *La probabilidad de que un auto no sufra siniestros por robo (total o parcial) en el próximo año es de .99. La función de densidad de la pérdida esta dada por:*

$$f_X(x) = 0.01\left[0.15e^{-0.15x}\right] \qquad x > 0 \text{ (en miles de pesos).}$$

El propietario del auto tiene una función de utilidad dada por

$$u(w) = -e^{-0.05w}.$$

Un asegurador ofrece al propietario un seguro que le pagará cualquier pérdida que pueda sufrir el auto el próximo año por concepto de robo (total o parcial) siempre y cuando exceda de 1 (mil pesos).

Calcula la máxima cantidad que el propietario del coche estará dispuesto a pagar como prima, para adquirir un seguro con las características mencionadas

Solución 2.4.4 *La utilidad por asegurar se plantea como sigue*

$$0.99u\left(w-G\right)+0.01\left[\int_0^1 u\left(w-X-G\right)f_X\left(x\right)dx+\int_1^\infty u\left(w-G\right)f_X\left(x\right)dx\right]$$

$$\begin{aligned}
&= -0.99e^{-0.05(w-G)}-.000015e^{-0.05(w-G)}\left[\int_0^1 e^{-0.05(-x)}e^{-0.15x}dx+\int_1^\infty e^{-0.15x}dx\right]\\
&= e^{-0.05(w-G)}\left[-0.99-.0000142743873-.000086079764\right]\\
&= -0.9901003452e^{-0.05(w-G)}.
\end{aligned}$$

y por asumir él solo el riesgo

$$0.99u\left(w\right)+0.01\int_0^\infty u\left(w-X\right)f_X\left(x\right)dx$$

$$\begin{aligned}
&= -0.99e^{-0.05w}-.000015e^{-0.05w}\int_0^\infty e^{-0.05(-x)}e^{-0.15x}dx\\
&= e^{-0.05w}\left[-0.99-.00015\right]\\
&= -.99015e^{-0.05w}.
\end{aligned}$$

Así que la solución resulta de igualar las dos funciones anteriores y despejar G.

$$G=.00100299.$$

Capítulo 3

Los procesos de decisión Markovianos

3.1 Introducción

Un problema de control óptimo consta de tres partes: un modelo matemático del sistema dinámico que se analiza, un conjunto de políticas o estrategias que es posible aplicar por un agente o tomador de decisiones y uno o más criterios de optimalidad para evaluar el efecto de las distintas políticas. En la sección 3.2 de este capítulo se expondrán estos tres elementos básicos para un modelo de decisión secuencial que evoluciona como se explicó en la sección 1.2. Debemos señalar que en esta breve exposición nos restringiremos a las estructuras matemáticas que requeriremos para el análisis y solución de los problemas de seguros que nos proponemos analizar. En las secciones 3.3 y 3.4 abordaremos una de las técnicas que se emplean para resolver un problema de decisión Markoviano, es decir, una técnica para encontrar una política óptima, técnica que además nos permite calcular el valor esperado de la recompensa que se obtendría al aplicar dicha política. Se trata del método de iteración de valores que será abordado para problemas que se observan en un número finito de épocas de decisión en la sección 3.3 y para problemas con un horizonte de planeación infinito en la sección 3.4.

3.2 Problemas de decisión Markovianos

3.2.1 Modelo matemático de un sistema aleatorio a tiempo discreto

El modelo del sistema se encuentra formado por los siguientes elementos:

T : El conjunto de épocas de decisión, el cual consideraremos discreto, es decir,

finito o numerable. Los elementos de este conjunto son los momentos en que el tomador de decisiones observa el estado en que se encuentra el sistema. Si este conjunto es finito, diremos que le modelo tiene horizonte de planeación finito; en otro caso, hablaremos de horizonte infinito. Los intervalos de tiempo que se forman entre dos épocas de decisión serán llamados períodos o etapas, y los supondremos de igual tamaño.

X : El conjunto de estados en los que puede encontrarse el sistema al ser observado. En este trabajo, los conjuntos de estados serán conjuntos discretos o subconjuntos compactos de un espacio Euclidiano.

A : El conjunto de acciones o controles que contiene las decisiones que se pueden tomar para afectar la evolución del sistema. Este conjunto también puede ser numerable o un subconjunto compacto de un espacio Euclidiano. No todas las acciones o controles son viables en todos los estados. Por ello, para cada estado x consideraremos un subconjunto de A, denotado por $\mathbf{A}_x$ o $\mathbf{A}(x)$, que representa las acciones viables cuando el sistema se encuentra en el estado x. De esta manera,

$$A = \bigcup_{x \in X} A_x.$$

$\mathbf{p}_t(x_{t+1} \,|x_t, a_t\,)$: Una probabilidad de transición que indica la probabilidad de que el sistema llegue al estado x_{t+1} dado que en la etapa anterior estaba en el estado x_t y que el tomador de decisiones eligió la acción a_t.

$\mathbf{r}_t(a, x)$: Una función recompensa que depende del estado en que se encuentra el sistema y de la acción elegida. Si esta función es negativa, se considera que se incurrió en un gasto en lugar de obtener una recompensa.

La quinteta formada por los elementos anteriores, es decir:

$$\{\mathbf{T}, \mathbf{X}, \mathbf{A}, \mathbf{r}\,(\mathbf{x}, \mathbf{a})\,, \mathbf{p}\,(\cdot \;|x, a)\}\,,$$

es el modelo matemático de un sistema dinámico aleatorio a tiempo discreto del tipo que nos interesa estudiar en este trabajo.

Algunas precisiones finales.

1. En cada uno de los espacios de estados y de acciones, se define una $\sigma-$álgebra. Supondremos que ésta siempre es la $\sigma-$álgebra de Borel (que en caso de que los espacios sean discretos, se reduce al conjunto potencia). Éstas $\sigma-$álgebras serán denotadas por $\mathfrak{B}\,(X)$ y $\mathfrak{B}\,(A)$.

2. Para facilitar la referencia a las parejas de estados y acciones que son viables, se acostumbra denotar este conjunto por

$$\mathbb{K} = \{(x, a)\,|x \in X, a \in A_x\}\,.$$

De esta manera, la función recompensa es una función $r_t : \mathbb{K} \to \mathbb{R}$ y la probabilidad de transición está condicionada por la ocurrencia de alguna pareja en $\mathbb{K}$.

3. En algunos modelos se encuentra que la recompensa, no sólo depende del estado en que se encuentra el sistema y de la acción que se elige, sino también del estado al que el sistema llegará en la siguiente época de decisión, es decir, se trata de una función

$$r_t : \mathbb{K} \times X \to \mathbb{R} \quad \text{donde } r(x_t, a_t, x_{t+1}) \in \mathbb{R}.$$

Para los cálculos posteriores, en estos casos tomamos la recompensa esperada dada por

$$r_t(x, a) = \sum_{j \in X} r_t(x, a, j)\, p_t(j \mid x, a).$$

4. Cuando el horizonte de planeación es finito, en $t = N$ no se elige ya ninguna acción de manera que la recompensa terminal es una función exclusivamente del último estado al que llega el sistema $r_N(x_N)$

3.2.2 Políticas de control

Una política o estrategia es una manera de asignar acciones en cada época de decisión, para cada estado posible del sistema. Cada política está formada por una colección de funciones $d_1, d_2, ...d_t, \ldots$ donde d_t indica la forma de elegir una acción en un momento particular de observación del sistema: el momento t. Estas funciones se conocen como *funciones de decisión* y pueden ser de diferentes tipos:

1. Si la regla sólo toma en cuenta el estado en que se encuentra el sistema en el momento de observación, tendremos una regla Markoviana. Cuando se elige una acción con certeza se habla de una regla determinista. Reuniendo las dos condiciones anteriores tenemos una regla Markoviana y determinista que se describe a través de una función del conjunto de estados al conjunto de acciones, es decir:

$$d_t : X \longrightarrow A, \quad \text{sujeta a la restricción} \quad d(x) \in A_x.$$

El conjunto de tales reglas de decisión será denotado por D_t^{MD}.

2. Cuando la función no determina una acción para cada estado sino una distribución de probabilidad sobre el conjunto de acciones viables, tenemos una regla de decisión Markoviana y aleatoria. Denotando el espacio de distribuciones de probabilidad en el conjunto A por $\mathcal{P}(A)$, tenemos:

$$d_t : X \to \mathcal{P}(A) \quad \text{donde} \quad d_t(x) = q_{d_t}(a) \quad \text{con} \quad a \in A_x,$$

siendo $q_{d_t}(\cdot)$ un elemento de $\mathcal{P}(A)$. Al conjunto de estas reglas de decisión lo denotaremos por D_t^{MR}.

3. Cuando la regla de decisión depende de la historia de estados y acciones ocurridos con anterioridad se dice que se trata de una regla dependiente de la historia. Denotamos la historia hasta el tiempo t como $h_t = (x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t)$ que a veces también se escribe recursivamente usando la historia hasta el tiempo $t-1$ como $h_t = (h_{t-1}, a_{t-1}, x_t)$. El conjunto de todas las historias hasta el tiempo t se denota por H_t. Entonces, una regla dependiente de la historia y determinista está dada por:

$$d_t : H_t \to A \quad \text{sujeta a la restricción} \quad d_t(h_t) \in A_{x_t}.$$

El conjunto de estas reglas será denotado por D_t^{HD}.

4. Una regla de decisión dependiente de la historia y aleatoria es una función que va del conjunto de historias al conjunto de distribuciones de probabilidad, es decir,

$$d_t : H_t \to \mathcal{P}(A) \quad \text{dada por} \quad d_t(h_t) = q_{d_t}(a) \quad \text{con} \quad a \in A_{x_t}.$$

Usaremos D_t^{HR} para designar al conjunto de estar reglas.

Las recompensas y las probabilidades de transición se convierten en funciones de x o h_t una vez que se ha determinado una regla de decisión.

En un modelo con horizonte finito que tenga N épocas de decisión, una política es de la forma $\pi = (d_1, \ldots, d_{N-1})$. Para horizonte infinito tendremos una sucesión infinita de reglas de decisión $\pi = (d_1, d_2, d_3, \ldots)$.

Si las reglas de decisión son Markovianas y deterministas, la política correspondiente lleva el mismo nombre. El conjunto de tales políticas será denotado por Π^{MD}. Análogamente para los casos Markoviano y aleatorio $\left(\Pi^{MR}\right)$, dependiente de la historia y determinista $\left(\Pi^{HD}\right)$ y dependiente de la historia y aleatorio $\left(\Pi^{HR}\right)$.

Hay un tipo particular de políticas que serán de mucha utilidad, llamadas políticas estacionarias, en las que usamos la misma regla de decisión en todas las épocas, es decir, $\pi = (d, d, \ldots)$. Para modelos de horizonte finito, las políticas estacionarias se denotan por $\pi = d^{N-1}$ y en el caso infinito por $\pi = d^{\infty}$. Denotamos Π^{SD} el conjunto de políticas estacionarias determinista y Π^{SR} el conjunto de políticas estacionarias aleatorizadas.

Evidentemente, las políticas más generales son las dependientes de la historia y aleatorias y las más particulares son las estacionarias y deterministas. Cuando no haya lugar a confusiones, usaremos simplemente Π para denotar el conjunto Π^{HR} que reúne a todas las políticas posibles (los otros conjuntos de políticas están contenidos en éste último). De igual manera, usaremos el símbolo D_t para referirnos a D_t^{HR}.

Ejemplo 3.2.1 *Para ejemplificar ahora los distintos tipos de reglas de decisión, asumamos que el conjunto de épocas de decisión es*

$$T = \{1, 2, 3\}.$$

donde tomamos decisiones en las épocas de decisión 1 y 2, y representamos a estas políticas por $\pi^K = \left(d_1^K, d_2^K\right)$ *,con* $K = MD, MR, HD,$ *o* HR*. Asumamos también que*

$$X = \{0, 1, 2, 3\},$$

y recordando que el valor máximo de $a \in A_x$ *es* $3-x$*,tenemos como una política posible en el caso* $K = MD$*:*

En la época de decisión 1

$d_1^{MD}(0) = 2$
$d_1^{MD}(1) = 2$
$d_1^{MD}(2) = 1$
$d_1^{MD}(3) = 0$

En la época de decisión 2

$d_2^{MD}(0) = 1$
$d_2^{MD}(1) = 0$
$d_2^{MD}(2) = 0$
$d_2^{MD}(3) = 0$

Si $K = MR$*, las probabilidades para las decisiones son:*

En la época de decisión 1

$$q_{d_1^{MR}(x)}(a)$$

x	$q_{d_1^{MR}(x)}(0)$	$q_{d_1^{MR}(x)}(1)$	$q_{d_1^{MR}(x)}(2)$	$q_{d_1^{MR}(x)}(3)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{3}$	$\frac{2}{3}$	0	$\times$
2	$\frac{3}{4}$	$\frac{1}{4}$	$\times$	$\times$
3	1	$\times$	$\times$	$\times$

En la época de decisión 2

$$q_{d_2^{MR}(x)}(a)$$

x	$q_{d_2^{MR}(x)}(0)$	$q_{d_2^{MR}(x)}(1)$	$q_{d_2^{MR}(x)}(2)$	$q_{d_2^{MR}(x)}(3)$
0	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
1	$\frac{1}{3}$	0	$\frac{2}{3}$	$\times$
2	$\frac{1}{2}$	$\frac{1}{2}$	$\times$	$\times$
3	1	$\times$	$\times$	$\times$

Si $K = HD$, una política posible es:

En la época de decisión 1, como la historia solo depende del estado anterior, podemos poner la misma tabla que en el caso Markoviano

$d_1^{HD}(0) = 2$
$d_1^{HD}(1) = 2$
$d_1^{HD}(2) = 1$
$d_1^{HD}(3) = 0$

En la época de decisión 2 y con base en la tabla anterior tenemos

(x,a)	$d_2^{HD}(x,a,0)$	$d_2^{HD}(x,a,1)$	$d_2^{HD}(x,a,2)$	$d_2^{HD}(x,a,3)$
$(0,2)$	3	2	0	$\times$
$(1,2)$	$\times$	1	0	0
$(2,1)$	$\times$	1	0	0
$(3,0)$	$\times$	1	1	0

Si $K = HR$,

En la época de decisión 1, como la historia solo depende del estado anterior, podemos poner la misma tabla que en el caso Markoviano

$$q_{d_1^{HR}(x)}(a)$$

x	$q_{d_1^{HR}(x)}(a)$	$q_{d_1^{HR}(x)}(a)$	$q_{d_1^{HR}(x)}(a)$	$q_{d_1^{HR}(x)}(a)$
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{3}$	$\frac{2}{3}$	0	$\times$
2	$\frac{3}{4}$	$\frac{1}{4}$	$\times$	$\times$
3	1	$\times$	$\times$	$\times$

En la época de decisión 2

(x,a)	$q_{d_2^{HR}(x,a,0)}(0)$	$q_{d_2^{HR}(x,a,0)}(1)$	$q_{d_2^{HR}(x,a,0)}(2)$	$q_{d_2^{HR}(x,a,0)}(3)$
$(0,0)$	0	$\frac{1}{2}$	$\frac{1}{2}$	0
$(0,1)$	$\frac{1}{4}$	$\frac{3}{4}$	0	0
$(0,2)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0
$(0,3)$	$\times$	$\times$	$\times$	$\times$
$(1,0)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$(1,1)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	0
$(1,2)$	$\times$	$\times$	$\times$	$\times$
$(2,0)$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
$(2,1)$	$\times$	$\times$	$\times$	$\times$
$(3,0)$	$\times$	$\times$	$\times$	$\times$

(x,a)	$q_{d_2^{HR}(x,a,1)}(0)$	$q_{d_2^{HR}(x,a,1)}(1)$	$q_{d_2^{HR}(x,a,1)}(2)$	$q_{d_2^{HR}(x,a,1)}(3)$
$(0,0)$	$\times$	$\times$	$\times$	$\times$
$(0,1)$	0	$\frac{1}{4}$	$\frac{3}{4}$	$\times$
$(0,2)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\times$
$(0,3)$	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$	$\times$
$(1,0)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\times$
$(1,1)$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\times$
$(1,2)$	0	$\frac{1}{4}$	$\frac{3}{4}$	$\times$
$(2,0)$	1	0	0	$\times$
$(2,1)$	0	1	0	$\times$
$(3,0)$	0	0	1	$\times$

,

(x,a)	$q_{d_2^{HR}(x,a,2)}(0)$	$q_{d_2^{HR}(x,a,2)}(1)$	$q_{d_2^{HR}(x,a,2)}(2)$	$q_{d_2^{HR}(x,a,2)}(3)$
$(0,0)$	$\times$	$\times$	$\times$	$\times$
$(0,1)$	$\times$	$\times$	$\times$	$\times$
$(0,2)$	$\frac{1}{4}$	$\frac{3}{4}$	$\times$	$\times$
$(0,3)$	$\frac{1}{2}$	$\frac{1}{2}$	$\times$	$\times$
$(1,0)$	$\times$	$\times$	$\times$	$\times$
$(1,1)$	$\frac{1}{5}$	$\frac{4}{5}$	$\times$	$\times$
$(1,2)$	$\frac{1}{4}$	$\frac{3}{4}$	$\times$	$\times$
$(2,0)$	1	0	$\times$	$\times$
$(2,1)$	0	1	$\times$	$\times$
$(3,0)$	0	1	$\times$	$\times$

,

(x,a)	$q_{d_2^{HR}(x,a,3)}(0)$	$q_{d_2^{HR}(x,a,3)}(1)$	$q_{d_2^{HR}(x,a,3)}(2)$	$q_{d_2^{HR}(x,a,3)}(3)$
$(0,0)$	$\times$	$\times$	$\times$	$\times$
$(0,1)$	$\times$	$\times$	$\times$	$\times$
$(0,2)$	$\times$	$\times$	$\times$	$\times$
$(0,3)$	1	$\times$	$\times$	$\times$
$(1,0)$	$\times$	$\times$	$\times$	$\times$
$(1,1)$	$\times$	$\times$	$\times$	$\times$
$(1,2)$	1	$\times$	$\times$	$\times$
$(2,0)$	$\times$	$\times$	$\times$	$\times$
$(2,1)$	1	$\times$	$\times$	$\times$
$(3,0)$	1	$\times$	$\times$	$\times$

.

3.2.3 Criterios de optimalidad

Nuestro propósito en esta sección es presentar distintas funciones que nos permitan evaluar el efecto de aplicar una u otra política. Es decir, se trata de definir criterios que nos permitirán establecer con claridad qué es lo que el tomador de decisiones considera un "buen funcionamiento" del sistema y buscar políticas óptimas de acuerdo a ese criterio.

Antes de que el sistema empiece a evolucionar, los estados y las acciones son variables aleatorias y las funciones recompensa $r_t(x, a)$ también lo son. Entonces los criterios de optimalidad estarán definidos en términos de funciones de los valores esperados de la sucesión de recompensas que se genera al aplicar una u otra política. A estas funciones se les conoce como *índices de funcionamiento.*

Inicialmente requerimos analizar en qué espacio de probabilidad se ubican las sucesiones de estados y acciones y, en consecuencia, las sucesiones de recompensas, para después hablar del operador esperanza correspondiente.

Procesos estocásticos

Para empezar, incluímos la definición de un proceso estocástico:

Definición 3.2.2 *Un* proceso estocástico *es una colección de variables aleatorias* $\{Y_t : t > 0\}$ *definidas sobre el mismo espacio muestral* Ω. *Si el proceso es* a tiempo discreto, *se trata de una sucesión de variables aleatorias* $\{Y_t : t \in \mathbb{N}\}$.

Es posible interpretar los valores que toma cada una de las variables como los estados en que se encuentra un sistema dinámico cuya evolución es aleatoria. Es decir, $Y_t = x$ significa que al tiempo t el sistema se encuentra en el estado x. Cuando el conjunto de estados es a lo más numerable, hablamos de una cadena estocástica. La transición a un nuevo estado j que será observado al tiempo $k+1$ sigue una ley de probabilidad dada por

$$\Pr\left[Y_{k+1} = j \,|\, Y_1 = x_1, Y_2 = x_2, \ldots Y_k = x_k\right],$$

que se conoce como probabilidad de transición. Cuando la transición a un nuevo estado depende de la historia ocurrida sólo a través del último estado, es decir, cuando

$$\Pr\left[Y_{k+1} = j \,|\, Y_1 = x_1, Y_2 = x_2, \ldots Y_k = x_k\right] = \Pr\left[Y_{k+1} = j \,|\, Y_k = x_k\right],$$

diremos que la cadena es de Markov.

Nótese que, a diferencia de los modelos de decisión secuencial, en este caso la evolución del sistema no está afectada por la acción de un agente externo sino que las probabilidades de transición dependen sólo de cuál fue el estado anterior.

Procesos estocásticos inducidos por un modelo de decisión secuencial

Para cada política $\pi \in \Pi$ denotamos por X_t la variable aleatoria que indica el estado del sistema al tiempo t y por Y_t la variable aleatoria que indica la acción elegida al tiempo t. Estas variables aleatorias generan los procesos estocásticos $\{X_t\}$, $\{Y_t\}$. Veamos cómo es el espacio de probabilidad donde están definidos estos procesos estocásticos.

Como primer elemento tenemos el espacio muestral Ω. Al considerar la evolución del sistema a lo largo de un horizonte finito, se genera un vector aleatorio $(X_1, Y_1, X_2, Y_2, \ldots, X_{N-1}, Y_{N-1}, X_N)$. En este caso el espacio muestral estará dado por $\Omega = X \times A \times \ldots \times X = (X \times A)^{N-1} \times X$. Para un modelo de horizonte infinito el vector aleatorio es $(X_1, Y_1, X_2, Y_2, \ldots)$ de donde $\Omega = X \times A \times \ldots = (X \times A)^{\infty}$.

Un elemento del espacio muestral $\omega \in \Omega$ es de la forma $\omega = (x_1, a_1, x_2, a_2, ..., x_N)$ con $N \leq \infty$; ω representa una posible evolución del sistema y la llamaremos trayectoria muestral. De acuerdo a la definición de las variables X_t y Y_t tenemos:

$$\begin{aligned} X_t(\omega) &= x_t \\ Y_t(\omega) &= a_t. \end{aligned}$$

El segundo elemento corresponde al espacio de eventos o a la $\sigma-$álgebra de conjuntos de Ω, donde cada evento es un subconjunto de Ω. Recordemos que en X se tiene la $\sigma-$ álgebra $\mathfrak{B}(X)$ y en A, $\mathfrak{B}(A)$. Es posible construir la $\sigma-$ álgebra producto $\mathcal{A}$ en Ω usando lo siguiente:

Definición 3.2.3 *Para cada $j = 1, 2, ...$, sea $(\Omega_j, \mathcal{A}_j)$ un espacio medible. Sea $\Omega = \prod_{j=1}^{\infty} \Omega_j$, el conjunto de todas las sucesiones $(\omega_1, \omega_2, ...)$ tal que $\omega_j \in \Omega_j$, $j = 1, 2, ...$ Un rectángulo en Ω es un conjunto $B^n = A_1 \times A_2 \times .... \times A_n$, donde $Aj \subset \Omega j$ para cada $j = 1, 2, , ..., n$. De este rectángulo definimos*

$$B_n = \{\omega \in \Omega : (\omega_1, ..., \omega_n) \in B^n\}.$$

El conjunto B_n es llamado el cilindro con base B^n; el cilindro es medible si el rectángulo lo es , i.e., si $B^n \in \prod_{j=1}^{n} \mathcal{A}_j$. La más pequeña $\sigma-$álgebra que contiene a los cilindros medibles se llama la $\sigma-$álgebra producto y la escribiremos como $\prod_{j=1}^{\infty} \mathcal{A}_j$. Aunque se use la notación $\prod_{j=1}^{n} \mathcal{A}_j$, esto no significa el producto cartesiano de las $\mathcal{A}_j$. El producto cartesiano es el conjunto de los rectángulos medibles, mientras la $\sigma-$álgebra producto es la mínima $\sigma-$álgebra que contiene a los rectángulos medibles.

La $\sigma-$álgebra $\mathcal{A}$ de Ω en nuestro modelo, es la $\sigma-$álgebra producto $\mathfrak{B}(X) \times \mathfrak{B}(A) \cdots \times \mathfrak{B}(A) \times \mathfrak{B}(X)$ en el caso de horizonte finito.

Como tercer elemento está la medida de probabilidad. Hemos hablado del proceso de estados $\{X_t\}$ y del proceso de acciones $\{Y_t\}$. Definiremos ahora el proceso de historias $\{Z_t\}$ dado por

$$Z_1(\omega) = x_1 \quad \text{y} \quad Z_t(\omega) = (x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t),$$

para cada trayectoria muestral $\omega \in \Omega$. Dada una política arbitraria π, es posible construir una medida de probabilidad P^π sobre el espacio medible $(\Omega, \mathcal{A})$ de tal forma que para toda $x \in X$

$$\begin{aligned} P^\pi[X_1 = x_1] &= 1, \\ P^\pi[Y_t = a_t | Z_t = h_t] &= q_{d_t(h_t)}(a), \\ P^\pi[X_{t+1} = j \mid Z_t = (h_{t-1}, a_{t-1}, x_t), Y_t = a_t] &= p_t(j \mid x_t, a_t). \end{aligned}$$

La existencia y unicidad de esta medida está garantizada por el Teorema de Ionescu Tulcea (Ash, p. 109) que no incluimos en este trabajo por que excede los límites teóricos del mismo.

Con esta medida de probabilidad P^π, la probabilidad de obtener una trayectoria muestral $\omega = (x_1, a_1, ..., x_N)$ al aplicar la política $\pi \in \Pi$, está dada por:

$$\begin{aligned} P^\pi(\omega) = {} & q_{d_1(x_1)}(a_1)\, p_1(x_2 \mid h_1, a_1)\, q_{d_2(h_2)}(a_2)\, p_2(x_3 \mid h_2, a_2) \\ & \cdots\, q_{d_{N-1}(h_{N-1})}(a_{N-1})\, p_{N-1}(x_N \mid h_{N-1}, a_{N-1}). \end{aligned} \tag{3.2.1}$$

A esa misma trayectoria muestral le corresponde una sucesión de recompensas $\{r_1(x_1, a_1), ..., r_{N-1}(x_{N-1}, a_{N-1}), r_N(x_N)\}$. Denotemos por $R_t = r_t(X_t, Y_t)$ la recompensa obtenida en el periodo $t < N$, y por $R_N = r_N(X_N)$ la recompensa terminal. Sea $R = (R_1, ..., R_N)$ una sucesión de recompensas y sea $\mathfrak{R}$ el conjunto de todas las posibles sucesiones de recompensas.

Una política π nos lleva a una distribución de probabilidad $P^\pi_{\mathfrak{R}}(\cdot)$ en $\mathfrak{R}$ de tal forma que la probabilidad de que las recompensas tomen los valores $\rho_1, \rho_2, ..., \rho_N$ es

$$\begin{aligned} & P^\pi_{\mathfrak{R}}(\rho_1, \rho_2, ..., \rho_N) \\ = {} & P^\pi[\{(x_1, a_1, ..., x_N) \mid (r_1(x_1, a_1), ..., r_{N-1}(x_{N-1}, a_{N-1}), r_N(x_N)) = (\rho_1, \rho_2, ..., \rho_N)\}]. \end{aligned}$$

Sea W una variable aleatoria definida en $(\Omega, \mathcal{A}, P^\pi)$. Definimos el valor esperado de W cuando se aplica una política π como:

$$E^\pi\{W\} = \sum_{\omega \in \Omega} W(\omega) P^\pi\{\omega\} = \sum_{w \in \mathbb{R}} w P^\pi\{\omega : W(\omega) = w\}.$$

Denotamos la esperanza condicional dado que el estado inicial es x por

$$E_x^{\pi}\{W\} = \sum_{z \in \mathbb{R}} zP(W = z \mid X_1 = x),$$

y la esperanza dado que la historia hasta el tiempo t es h_t por

$$E_{h_t}^{\pi}[W] = \sum zP[W = z \mid Z_t = h_t].$$

Observación 3.2.4 *(a) Si W es función de $x_t, a_t, ..., a_{N-1}, x_N$; y $h_t \in H_t$ entonces*

$$E_x^{\pi}(W(X_t, Y_t, ..., X_{N-1}, Y_N)) = \sum W(x_t, a_t, ..., a_N, x_N) P^{\pi}(a_t, x_{t+1}, ..., a_{T-1}, x_N \mid x_t);$$

(b) Si v es una función de x_{t+1} entonces

$$E_x^{\pi}(v(x_{t+1} \mid h_t)) = \sum_{y \in S} v(y) p_t(y \mid x_t, d_t(h_t)). \tag{3.2.2}$$

Si W es una variable aleatoria absolutamente continua, estos valores esperados se calculan con integrales en vez de sumas.

Comparación de vectores aleatorios

Como requeriremos evaluar el efecto de distintas políticas, necesitamos definir ordenamientos estocásticos entre variables aleatorias para poder decir cual política preferimos. Decimos que la variable aleatoria U es estocásticamente mayor que la variable aleatoria V si

$$P(V > t) \leq P(U > t),$$

para toda $t \in \mathbb{R}$. Si P^1 y P^2 son dos distribuciones de probabilidad en el mismo espacio de probabilidad, entonces P^1 es estocásticamente mayor que P^2 si

$$P^2[(t, \infty)] \leq P^1[(t, \infty)],$$

para toda $t < \infty$.

En cuanto a vectores aleatorios, se han propuesto muchas generalizaciones, de las cuales, la más adecuada es la siguiente. Decimos que el vector aleatorio $\mathbf{U} = (U_1, ..., U_n)$ es estocásticamente mayor que el vector $\mathbf{V} = (V_1, ..., V_n)$ si

$$E[f(V_1, ..., V_n)] \leq E[f(U_1, ..., U_n)],$$

para toda $f : \mathbb{R}^n \to \mathbb{R}$, para la cual la esperanza exista y que preserve un orden parcial en $\mathbb{R}^n$ (es decir, que si $v_i \leq u_i$ para $i = 1, 2, ..., N$, entonces $f(V_1, ..., V_n) \leq f(U_1, ..., U_n)$).

Proposición 3.2.5 *Supongamos que* $\mathbf{U} = (U_1, ..., U_n)$ *y* $\mathbf{V} = (V_1, ..., V_n)$ *son vectores aleatorios tales que* $P(V_1 > t) \leq P(U_1 > t)$ *para toda* $t \in \mathbb{R}$; *y para* $j = 2, ..., N$ *si se cumple que* $vi \leq u_i$ *para* $i = 1, ..., j-1$, *entonces*

$$P[V_j > t \mid V_1 = v_1, ..., V_{j-1} = v_{j-1}] \leq P[U_j > t \mid U_1 = u_1, ..., U_{j-1} = u_{j-1}].$$

En este caso diremos que $\mathbf{U}$ *es estocásticamente mayor que* $\mathbf{V}$. *Además, para cualquier función* $g : \mathbb{R} \to \mathbb{R}$ *no decreciente.*

$$E[g(V_j) | V_1 = v_1, ..., V_{j-1} = v_{j-1}] \leq E[g(U_j) > t | U_1 = u_1, ..., U_{j-1} = u_{j-1}].$$

La dificultad de usar un ordenamiento estocástico para comparar políticas es que para que el tomador de decisiones prefiera π a ν, se debe cumplir la desigualdad

$$E^{\pi}[f(R_1, ..., R_n)] \geq E^{v}[f(R_1, ..., R_n)] \tag{3.2.3}$$

para una clase muy grande de funciones f, muchas de las cuales no reflejan la actitud del tomador de decisiones frente al riesgo. La teoría de utilidad brinda una alternativa para comparar políticas Como habíamos mencionado, la utilidad que denotaremos por $u(\cdot)$, es una función de valores reales que refleja las preferencias del tomador de decisiones en un conjunto dado. Para funciones de utilidad lineales y aditivas (es decir, $u(a_1, ..., a_n) = \sum_{i=1}^{n} a_i$) es fácil verificar que se cumple la desigualdad (3.2.3) a diferencia de otro tipo de funciones. Si el tomador de decisiones no prefiere v sobre w, entonces $u(v) \leq u(w)$, y si es indiferente, entonces $u(v) = u(w)$.

Para un vector aleatorio discreto Y, la utilidad esperada sobre el conjunto de preferencias W, esta dada por

$$E(u(Y)) = \sum_{y \in W} u(y) P[Y = y],$$

y para un vector aleatorio absolutamente continuo

$$E(u(Y)) = \int_{W} u(y) dF(y).$$

En un proceso de decisión Markoviano con espacio de estados discreto, la utilidad esperada al aplicar la política π se puede representar como

$$E^{\pi}[u(R)] = \sum_{(\rho_1, ..., \rho_n) \in \Re} u(\rho_1, ..., \rho_n) P^{\pi}_{\Re}\{(\rho_1, ..., \rho_n)\},$$

donde $\Re$ representa el conjunto de realizaciones de la sucesión de recompensas.

Bajo el criterio de utilidad esperada, el tomador de decisiones prefiere la política π a la polítca v si

$$E^{\pi}[u(R_1, ..., R_n)] > E^{v}[u(R_1, ..., R_n)].$$

Los índices de funcionamiento

De lo anterior se desprende que los cálculos se reducen si la función utilidad es lineal. Esta es la razón por la cual los índices de funcionamiento más comúnmente usados son precisamente de este tipo. A continuación describimos los tres tipos de índices que con mayor frecuencia se encuentran en la literatura sobre proceso de decisión Markovianos:

Recompensa total esperada Para horizonte finito se define

$$v_N^{\pi}(x) = E_x^{\pi}\left\{\sum_{t=1}^{N-1} r_t(X_t, Y_t) + r_N(X_N)\right\},$$

donde x es el estado inicial y r_N es la recompensa terminal.

Para horizonte infinito la definición será

$$v^{\pi}(x) = E_x^{\pi}\left\{\sum_{t=1}^{\infty} r_t(X_t, Y_t)\right\}.$$

El principal problema al trabajar con este índice es que es muy fácil que la serie sea divergente. El siguiente índice resuelve mejor este problema.

Recompensa descontada esperada En este índice cada recompensa se multiplica por una potencia de un factor de descuento $\lambda \in (0,1)$.

$$v_{\lambda}^{\pi}(x) = E_x^{\pi}\left\{\sum_{t=1}^{\infty} \lambda^{t-1} r(X_t, Y_t)\right\}.$$

Como $\sum_k \lambda^k$ converge geométricamente, basta con que las recompensas sean acotados para tener un serie convergente.

Este índice se aplica en problemas en donde las recompensas son cantidades monetarias, pues el factor de descuento puede interpretarse como el factor de valor presente de la recompensa que se obtendrá en un tiempo futuro t. Si δ representa la tasa de interés en el mercado, al cabo de un período un capital D se transformará en $C = (1+\delta)\, D$ de donde $D = \left(\frac{1}{1+\delta}\right) C$. Así $\lambda = \frac{1}{1+\delta}$ es el factor por el que debemos multiplicar C para calcular su valor presente cuando se aplica la tasa de interés δ.

Políticas Optimas

Una política π^* que satisface que

$$v^{\pi^*}(x) = \sup_{\pi \in \Pi} v^{\pi}(x),$$

es una política recompensa–total óptima. Análogamente, una π^* es recompensa–descontada óptima si

$$v_\lambda^{\pi^*}(x) = \sup_{\pi \in \Pi} v_\lambda^{\pi}(x).$$

Evidentemente, en caso de existir políticas óptimas, los supremos se convierten en máximos. Si no existen políticas que maximicen el índice de funcionamiento requerido, es posible trabajar con políticas ε–óptimas, es decir, políticas que difieren del supremo en menos que un ε seleccionado de antemano. Precisemos: una política π^* es ε–óptima de acuerdo a cualquier índice de funcionamiento $I(x)$, si satisface la desigualdad

$$I^{\pi^*}(x) + \varepsilon > \sup_{\pi \in \Pi} I^{\pi}(x).$$

El problema es entonces encontrar políticas óptimas o ε-óptimas y calcular el valor de la recompensa esperada (total o descontada) que se obtendría al aplicar dichas políticas óptimas.

Un *problema de control Markoviano* consiste de un modelo de control Markoviano $\{T, X, A, r(x,a), p(\cdot \mid x,a)\}$, y de un índice de funcionamiento. Dado este problema su resolución consiste en encontrar una política bajo la cual el índice de funcionamiento alcanza su valor óptimo.

3.3 Problemas con horizonte finito y el algoritmo de programación dinámica para recompensa descontada.

Encontrar una política óptima para un problema que consta de un solo período, resulta muy sencillo. En este caso, se tendría una época de decisión $t = 1$ en la cual el sistema está en un estado x y el agente elige una acción a. En ese momento, se recibe una recompensa $r_1(x,a)$ y el sistema evoluciona hacia un estado terminal j con probabilidad:

$$p(j \mid x,a).$$

Así, si el espacio de estados X es discreto, la recompensa terminal esperada será

$$E_x^{\pi}[r_2(X_2)] = \sum_{j \in X} r_2(j)\, p(j \mid x,a),$$

donde π es una política formada exclusivamente por una regla de decisión, $\pi = (d)$ que cumple $d(x) = a \in A_x$. La recompensa total esperada es entonces

$$v_2^d(x) = r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x,a).$$

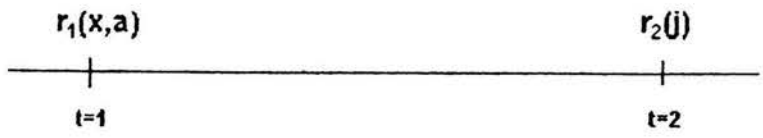

Figure 3.1: Recompensa inicial y final para un sólo período.

Pero si antes de aplicar política alguna queremos saber cuál es la mejor, buscamos a^* que satisfaga que

$$r_1(x, d^*(x)) + \sum_{j \in X} r_2(j)\, p(j \mid x, d^*(x)) = \max_{a \in A_x} \left\{ r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right\}. \tag{3.3.4}$$

Al conjunto de acciones que maximizan lo denotamos por

$$\arg\max_{a \in A_x} \left\{ r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right\}.$$

Este conjunto puede estar formado por más de una acción. Si a^* está en el conjunto anterior, una política óptima es de la forma $\pi^* = (d)$ donde $d(x) = a^*$. Como aquí la historia se reduce al estado inicial x, nuestra política es dependiente de la historia y determinista. Cabe la pregunta de si es posible obtener una recompensa total esperada mayor seleccionando aleatoriamente la acción. Si existe el máximo del lado derecho de la ecuación (3.3.4), la recompensa total esperada no se puede mejorar. Para fundamentar esta afirmación recordemos que la recompensa total esperada para una apolítica $\pi \in \Pi^{HR}$ está dada por

$$\sum_{a \in A_x} q_d(a) \left[r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right],$$

donde $q_d \in \mathcal{P}(A)$ satisface que $q_d(a) > 0$ para toda acción $a \in A_x$ y $\sum_a q_d(a) = 1$. El máximo obtenido en (3.3.4) satisface, para toda $q_d \in \mathcal{P}(A)$,

$$\begin{aligned} \max_{a \in A_x} \left\{ r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right\} &= \sum_{a \in A_x} q_d(a) \left[\max_{a \in A_x} \left\{ r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right\} \right] \\ &\geq \sum_{a \in A_x} q_d(a) \left[r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right], \end{aligned}$$

de donde se desprende que

$$\max_{a \in A_x} \left\{ r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right\} \geq \max_{q_d \in \mathcal{P}(A)} \left\{ \sum_{a \in A_x} q_d(a) \left[r_1(x,a) + \sum_{j \in X} r_2(j)\, p(j \mid x, a) \right] \right\}.$$

En el caso en que el espacio de estados sea un subconjunto compacto de $\mathbb{R}$ y la probabilidad de transición sea una densidad de transición, la sumatoria que hemos usado en las expresiones anteriores deberá sustituirse por la integral

$$\int_X r_2(j)\, p(j \mid x, a)\, dj.$$

Para facilitar la exposición, en lo que sigue de este capítulo consideraremos sólo el caso de un espacio de estados discreto. Todas las expresiones que obtengamos pueden fácilmente transformarse para el caso de un espacio de estados compacto.

3.3.1 El algoritmo de programación dinámica

Resolver un problema de decisión Markoviano con $N-1$ períodos (N épocas de decisión), no es tan sencillo. Dada una política π, tenemos una distribución de probabilidades sobre el conjunto de las historias. Para cada historia hay una sucesión de recompensas. Antes de que el sistema empiece a evolucionar, las recompensas son funciones de las variables aleatorias $X_t(h_t)$ y $Y_t(h_t)$, de manera que

$$\mathbf{R} = (r(X_1, Y_1), \cdots, r(X_{N-1}, Y_{N-1}), r(X_N))$$

es un vector aleatorio con una densidad conjunta que se desprende de las distribuciones de probabilidad de las historias (ver la ecuación (3.2.1)). Por lo tanto, para calcular el valor esperado

$$v_N^\pi(x) = E_x^\pi \left\{ \sum_{t=1}^{N-1} r_t(X_t, Y_t) + r_N(X_N) \right\}, \tag{3.3.5}$$

requeriríamos conocer la probabilidad de que $\mathbf{R}$ tome el valor $(\rho_1, \rho_2, \ldots \rho_N)$ para cualquier $N-$ada, es decir, conocer

$$P^\pi(\rho_1, \rho_2, \ldots \rho_N) = P^\pi\left[\{(x_1, a_1, \ldots, x_{N-1}, a_{N-1}, x_N) \mid r(x_t, a_t) = \rho_t, t = 1, \ldots, N\}\right],$$

para cualquier vector $(\rho_1, \rho_2, \ldots \rho_N) \in \mathbb{R}^n$, lo que resulta prácticamente imposible.

El algoritmo de programación dinámica le da la vuelta a este problema aplicando un mecanismo "de atrás para adelante" que permite transformar un problema de N períodos en N problemas de un solo período.

Veamos primero cómo calcular la recompensa total esperada (3.3.5) para una política $\pi = (d_1, d_2, \ldots, d_{N-1})$ dada. La idea es empezar en el último período y calcular

$$w_{N-1}^\pi(x) = r_{N-1}(x, d_{N-1}) + \sum_{j \in X} r_N(j)\, p(j \mid x, d_{N-1});$$

luego, utilizar $w_{N-1}^\pi(x)$ como recompensa terminal en el período anterior y calcular

$$w_{N-2}^\pi(x) = r_{N-2}(x, d_{N-1}) + \sum_{j \in X} w_{N-1}^\pi(j)\, p(j \mid x, d_{N-2}),$$

y así sucesivamente. Mostraremos que siguiendo este procedimiento, la función que se obtiene en el último paso es precisamente la recompensa total esperada.

El algoritmo de evaluación de la R.T.E. :

Sea $\pi = (d_1, d_2, \ldots d_{N-1})$ una política determinista histórico-dependiente

1. Tomar como función w_N^π a la recompensa terminal:

$$w_N^\pi(h_N) = r_N(x_N),$$

para cada historia h_N que termine en x_N, es decir, $h_N = (h_{N-1}, a_{N-1}, x_N)$.

2. Calcular $w_t^\pi(h_t)$ para cada historia h_t que termine en $x_t \in X$, mediante

$$\begin{aligned} w_t^\pi(h_t) &= r_t(x_t, d_t(h_t)) + \sum_{j \in X} w_{t+1}^\pi(h_t, d_t(h_t), j)\, p(j \mid x_t, d_t(h_t)) \quad (3.3.6) \\ &= r_t(x_t, d_t(h_t)) + E_{h_t}^\pi \left[w_{t+1}^\pi(h_t, d_t(h_t), X_{t+1}) \right]. \quad (3.3.7) \end{aligned}$$

3. Si $t = 1$, parar. De otra manera, reemplazar t por $t-1$ y regresar al paso 2.

Ahora queremos probar que $w_1^\pi(x)$ es precisamente la recompensa total esperada al aplicar la política π dado que el estado inicial es x.

Proposición 3.3.1 *La función $w_1^\pi(x)$ obtenida en algoritmo anterior cumple*

$$w_1^\pi(x) = v_N^\pi(x) = E_x^\pi \left\{ \sum_{t=1}^{N-1} r_t(X_t, Y_t) + r_N(X_N) \right\}. \quad (3.3.8)$$

Demostración. Demostraremos que en cada paso del algoritmo, la función $w_t^\pi(h_t)$ que se obtiene, es precisamente la recompensa total esperada al aplicar π del tiempo t en adelante suponiendo que ha ocurrido la historia h_t, es decir,

$$w_t^\pi(h_t) = E_{h_t}^\pi \left[\sum_{n=t}^{N-1} r_n(X_n, Y_n) + r_N(X_N) \right]. \quad (3.3.9)$$

Es claro que, si se cumple la igualdad anterior para toda t, en particular para $t = 1$ se tiene la conclusión deseada.

Probaremos (3.3.9) por inducción hacia atrás: el resultado es obvio para $t = N$. Supongamos que se cumple para $t+1, t+2, \ldots N$. Entonces, usando la expresión (3.3.7)

$$\begin{aligned} w_t^\pi(h_t) &= r_t(x_t, d_t(h_t)) + E_{h_t}^\pi \left[w_{t+1}^\pi(h_t, d_t(h_t), X_{t+1}) \right] \\ &= r_t(x_t, d_t(h_t)) + E_{h_t}^\pi \left[E_{h_{t+1}}^\pi \left[\sum_{n=t+1}^{N-1} r_n(X_n, Y_n) + r_N(X_N) \right] \right] \\ &= r_t(x_t, d_t(h_t)) + E_{h_t}^\pi \left[\sum_{n=t+1}^{N-1} r_n(X_n, Y_n) + r_N(X_N) \right] \\ &= E_{h_t}^\pi \left[\sum_{n=t}^{N-1} r_n(X_n, Y_n) + r_N(X_N) \right], \end{aligned}$$

donde la última igualdad se debe a que el primer sumando puede incluirse en la esperanza debido a que al tiempo t, el estado x_t y la historia h_t se conocen. ■

Si se quisiera aplicar el algoritmo anterior a una política aleatorizada en lugar de una determinista, se deben reemplazar las ecuaciones de la forma (3.3.6) por

$$w_t^{\pi}(h_t) = \sum q_{d_t(h_t)}(a) \left\{ r_t(x_t, d_t(h_t)) + \sum_{j \in X} w_{t+1}^{\pi}(h_t, d_t(h_t), j)\, p(j \mid x_t, d_t(h_t)) \right\},$$

y se puede demostrar que también en este caso se cumple la igualdad (3.3.9), es decir, que se tiene el mismo resultado dado por la proposición anterior para políticas aleatorizadas.

Supongamos ahora que queremos determinar la política π^* que maximice la recompensa total esperada.

El algoritmo de programación dinámica:

1. Se fija $t = N$ y se parte de la condición inicial

$$w_N^*(h_N) = r_N(x_N)$$

para toda historia h_N que termine en x_N.

2. Calculamos $w_t^*(h_t)$ para cada historia h_t que termine en $x_t \in X$, mediante

$$w_t^*(h_t) = \max_{a \in A_{x_t}} \left\{ r_t(x_t, a) + \sum_{j \in X} w_{t+1}^*(h_t, a, j)\, p_t(j \mid x_t, a) \right\},$$

y elegimos una acción $d_t^*(h_t)$ en el conjunto

$$A_{x_t}^* = \arg\max_{a \in A_{x_t}} \left\{ r_t(x_t, a) + \sum_{j \in X} w_{t+1}^*(h_t, a, j)\, p_t(j \mid x_t, a) \right\}.$$

3. Si $t = 1$, paramos. De otra manera, reemplazamos t por $t - 1$ y regresamos al paso 2.

Teorema 3.3.2 *Si para cada $t = N, N-1, \ldots, 1$, existe $d_t^* \in D$ que a cada estado x_t le asigna una acción en el conjunto de maximizadores A_{x_t}, es decir, tal que*

$$w_t^*(h_t) = r_t(x_t, d_t^*(x_t)) + \sum_{j \in X} w_{t+1}^*(h_t, a, j) p_t(j \mid x_t, d_t^*(x_t)) \qquad \forall\, x_t \in X.$$

Entonces

$$(i)\ \ w_1^*(x) = \sup_{\pi \in \Pi} v_{N,\lambda}^{\pi}(x) = \sup_{\pi \in \Pi} E_x^{\pi} \left[\sum_{t=1}^{N-1} r_t(x_t, a_t) + r_N(x_N) \right].$$

(*ii*) *La política* $\pi^* = \{d_1^*, d_2^*, \ldots\ldots, d_{N-1}^*)$ *es óptima, es decir,* $v_{N,\lambda}^{\pi^*}(x) = \sup_{\pi\in\Pi} v_{N,\lambda}^{\pi}(x)$.

Demostración. (i) Probaremos que cada una de las funciones w_t^* definidas en el algoritmo, cumplen la igualdad

$$w_t^*(h_t) = \sup_{\pi\in\Pi} E_{h_t}^{\pi}\left[\sum_{n=t}^{N-1} r_n(X_n, Y_n) + r_N(X_N)\right],$$

es decir,

$$w_t^*(h_t) = \sup_{\pi\in\Pi} w_t^{\pi}(h_t).$$

En tal caso, resulta obvio que se cumpliría

$$w_1^*(h_1) = w_1^*(x) = \sup_{\pi\in\Pi} v_{N,\lambda}^{\pi}(x).$$

Primero tenemos que mostrar que $w_t^*(h_t)$ es cota superior del conjunto

$$\{w_t^{\pi}(h_t) \mid \pi \in \Pi\}.$$

. Sea $\pi = (d_1, d_2, \ldots, d_{N-1})$ una política arbitraria en Π. Es claro que $w_N^*(h_N) = r_N(x_N) = w_N^{\pi}(h_N)$ para toda historia h_N que acabe en x_N. Supongamos que $w_t^*(h_t) \geq w_t^{\pi}(h_t)$ para $t+1, \ldots, N-1, N$. Entonces, para t tenemos

$$\begin{aligned}
w_t^*(h_t) &= \max_{a\in A_{x_t}}\left\{r_t(x_t,a) + \sum_{j\in X} w_{t+1}^*(h_t,a,j)\, p_t(j \mid x_t,a)\right\} \\
&\geq \max_{a\in A_{x_t}}\left\{r_t(x_t,a) + \sum_{j\in X} w_{t+1}^{\pi}(h_t,a,j)\, p_t(j \mid x_t,a)\right\} \\
&\geq \sum q_{d(h_t)}(a)\left\{r_t(x_t,a) + \sum_{j\in X} w_{t+1}^{\pi}(h_t,a,j)\, p_t(j \mid x_t,a)\right\} \\
&= w_t^{\pi}(h_t).
\end{aligned}$$

Ahora veamos que es la mínima cota superior. Sea $\varepsilon > 0$. Construyamos una política $\pi = (d_1, d_2, \ldots, d_{N-1})$ eligiendo $d_t(h_t)$ que satisfaga

$$r_t(x_t, d_t^*(h_t)) + \sum_{j\in X} w_{t+1}^*(x_t, d_t(h_t), j)\, p_t(j \mid x_t, d_t(x_t)) + \varepsilon \geq w_t^*(h_t).$$

Esa política satisface que

$$\begin{aligned}
w_t^{\pi}(h_t) &= r_t(x_t, d_t(h_t)) + \sum_{j\in X} w_{t+1}^{\pi}(x_t, d_t(h_t), j)\, p_t(j \mid x_t, d_t(x_t)) \\
&\geq r_t(x_t, d_t(h_t)) + \sum_{j\in X} w_{t+1}^*(x_t, d_t(h_t), j)\, p_t(j \mid x_t, d_t(x_t)) - (N-t-1)\,\varepsilon \\
&\geq w_t^*(h_t) - (N-t-1)\,\varepsilon.
\end{aligned}$$

(ii) Tomemos ahora una política $\pi^* = \{d_1^*, d_2^*, ..., d_{N-1}^*)$. De lo anterior es claro que, para toda t,

$$w_t^{\pi^*}(h_t) = w_t^*(h_t).$$

En particular, para $t = 1$, de donde se desprende la conclusión buscada. ■

Observación 3.3.3 *1. El algoritmo de programación dinámica no sólo conduce a políticas óptimas desde la primera época de decisión sino que brinda políticas que son óptimas aun cuando se inicie el cálculo en cualquier época de decisión t dado que ha ocurrido cualquiera de las historias que pueden ocurrir hasta ese momento.*

2. *La política óptima que se obtiene mediante este algoritmo, es* ***determinista y Markoviana.***

3. *Si el espacio de acciones es finito, siempre existen acciones maximizadoras en cada paso. Si el espacio de acciones es un compacto, se cumple 3.4.1, y las funciones recompensa son continuas respecto a las acciones, asi como también las probabilidades de transición, entonces existe una política Markoviana determinista que es óptima. Esto lo vemos de la siguiente manera, como la suma de dos funciones continuas es continua también,* $r_t(x_t, a) + \sum_{j \in X} p_t(j \mid x_t, a)$ *es continua respecto a las acciones y como culalquier función continua sobre un compacto, a su vez es un compacto, la expresión anterior también lo es y por lo tanto esta acotada. Como* $0 \leq r_t(x, a) \leq M$ *para toda* $x \in X, a \in A$, *entonces* $0 \leq w_t(j) \leq NM$ *para toda* $x \in X$, *por lo que* $r_t(x_t, a) + \sum_{j \in X} w_{t+1}(j) p_t(j \mid x_t, a) \leq r_t(x_t, a) + NM \sum_{j \in X} p_t(j \mid x_t, a)$ *para toda* $x \in X, a \in A$ *y* $t = 1, 2, ..., N$. *Entonces existe* a^* *tal que* $r_t(x_t, a^*) + \sum_{j \in X} w_{t+1}(j) p_t(j \mid x_t, a^*)$ *alcanza el máximo.*

De las observaciones anteriores, se desprende que podemos escribir el algoritmo de programación dinámica de la siguiente manera:

Algoritmo de programación dinámica Markoviano

1. Se fija $t = N$ y se parte de la condición inicial

$$w_N^*(x_N) = r_N(x_N),$$

para cada estado terminal x_N.

2. Calculamos $w_t^*(x_t)$ para cada estado $x_t \in X$, mediante

$$w_t^*(x_t) = \max_{a \in A_{x_t}} \left\{ r_t(x_t, a) + \sum_{j \in X} w_{t+1}^*(j) p_t(j \mid x_t, a) \right\},$$

y elegimos una acción $d_t^*(x_t)$ en el conjunto

$$A_{x_t}^* = \arg\max_{a \in A_{x_t}} \left\{ r_t(x_t, a) + \sum_{j \in X} w_{t+1}^*(j)\, p_t(j \mid x_t, a) \right\}.$$

3. Si $t = 1$, paramos. De otra manera, reemplazamos t por $t-1$ y regresamos al paso 2.

Para ejemplificar el algoritmo retomaremos nuestro ejemplo de inventarios para $N = 4$.

Ejemplo 3.3.4 *Sea* $k = 4$, $c(u) = 2u, h(u) = u, l(u) = 0, M = 3, f(u) = 8u, g(u) = r_2(u) = 0$ *y,*

$$p_j = \begin{cases} \frac{1}{4} \; si \; j = 0 \\ \frac{1}{2} \; si \; j = 1 \\ \frac{1}{4} \; si \; j = 2 \end{cases}.$$

El modelo se interpreta como sigue. El inventario esta limitado a ser de 3 unidades. Todos los costos y ganancias son lineales. La recompensa o ganancia esperada, cuando se tienen $u = x + a$ *unidades de almacenamiento antes de cualquier pedido, esta dada por*

u	$F(u) = \sum_{j=0}^{u-1} f(j)\, p_j + f(u) \sum_{j=u}^{\infty} p_j$
0	0
1	$0 * \frac{1}{4} + 8 * \frac{3}{4} = 6$
2	$0 * \frac{1}{4} + 8 * \frac{1}{2} + 16 * \frac{1}{4} = 8$
3	$0 * \frac{1}{4} + 8 * \frac{1}{2} + 16 * \frac{1}{4} + 16 * 0 = 8$

La recompensa o ganancia después de costos esta dada por

$$r_t(x, a) = F(u) - O(a) - h(u) = \begin{cases} F(u) - x \; si \; a = 0 \\ F(u) - (4 + x + 3a) \; si \; a > 0 \end{cases},$$

de tal manera que queda

$r_t(x, a)$

x	$a=0$	$a=1$	$a=2$	$a=3$
0	0	-1	-2	-5
1	5	0	-3	$\times$
2	6	-1	$\times$	$\times$
3	5	$\times$	$\times$	$\times$

y recordando que

$$p_t(j \mid x,a) \begin{cases} 0 & si \quad M \geq j > x+a \\ p_{x+a-j} & si \quad M \geq x+a \geq j > 0 \\ \sum_{j=x+a}^{\infty} p_j & si \quad M \geq x+a \ y \ j = 0 \end{cases},$$

$p_t(j \mid x,a)$

$x+a$	$j=0$	$j=1$	$j=2$	$j=3$
0	1	0	0	0
1	$\frac{3}{4}$	$\frac{1}{4}$	0	0
2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0
3	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Sea $N = 4$, *entonces,*

1.- Se fija $t = 4$ *y* $w_4^*(x_t) = r_4(x) = 0$, $x = 0,1,2,3$.

2.- Como $t \neq 1$, *continuamos. Fijamos* $t = 3$ *y*

$$w_3^*(x) = \max_{a \in A_x} \left\{ r(x,a) + \sum p(j \mid x,a)\, w_4^*(j) \right\}$$

$$= \max_{a \in A_x} \{ r(x,a) \}, \quad x = 0,1,2,3.$$

Observando los valores de $r(x,a)$ *vemos que la acción que maximiza la ganancia después de costos es*

x	$w_3^*(x)$	$A_{x,3}^*$
0	0	0
1	5	0
2	6	0
3	5	0

3.- Como $t \neq 1$ *continuamos. Fijamos* $t = 2$ *y*

$$w_2^*(x) = \max_{a \in A_x} \{ w_2^*(x,a) \},$$

donde, por ejemplo,

$$\begin{aligned} w_2^*(0,2) &= r(0,2) + p(0 \mid 0,2)\, w_3^*(0) + p(1 \mid 0,2)\, w_3^*(1) + p(2 \mid 0,2)\, w_3^*(2) + p(3 \mid 0,2)\, w_3^*(3) \\ &= -2 + \frac{1}{4} * 0 + \frac{1}{2} * 5 + \frac{1}{4} * 6 + 0 * 5 = 2. \end{aligned}$$

Las cantidades $w_2^*(x,a)$, $w_2^*(x)$ *y* $A_{x,2}^*$ *quedan resumidas en la siguiente tabla, con* $\times's$ *como acciones inalcanzables*

$$w_2^*(x,a)$$

x	$a=0$	$a=1$	$a=2$	$a=3$	$w_2^*(x)$	$A_{x,2}^*$
0	0	$\frac{1}{4}$	2	$\frac{1}{2}$	2	2
1	$\frac{25}{4}$	4	$\frac{5}{2}$	$\times$	$\frac{25}{4}$	0
2	10	$\frac{9}{2}$	$\times$	$\times$	10	0
3	$\frac{21}{2}$	$\times$	$\times$	$\times$	$\frac{21}{2}$	0

4.- Como $t \neq 1$ *continuamos. Fijamos* $t = 1$ *y*

$$w_1^*(x) = \max_{a \in A_x} \{w_1^*(x,a)\}.$$

Las cantidades $w_1^*(x,a), w_1^*(x)$ *y* $A_{x,1}^*$ *quedan resumidas en la siguiente tabla,*

$w_1^*(x,a)$

x	$a=0$	$a=1$	$a=2$	$a=3$	$w_1^*(x)$	$A_{x,2}^*$
0	2	$\frac{33}{16}$	$\frac{66}{16}$	$\frac{67}{16}$	$\frac{67}{16}$	3
1	$\frac{129}{16}$	$\frac{98}{16}$	$\frac{99}{16}$	$\times$	$\frac{129}{16}$	0
2	$\frac{194}{16}$	$\frac{131}{16}$	$\times$	$\times$	$\frac{194}{16}$	0
3	$\frac{227}{16}$	$\times$	$\times$	$\times$	$\frac{227}{16}$	0

5.- Como $t = 1$, *paramos.*

Este algoritmo produce la función de recompensa total esperada $w_3^*(x)$ *y la política óptima* $\pi^* = (d_1^*(x), d_2^*(x), d_3^*(x))$, *tabulada a continuación.*

x	$d_1^*(x)$	$d_2^*(x)$	$d_3^*(x)$	$w_3^*(x)$
0	3	2	0	$\frac{67}{16}$
1	0	0	0	$\frac{129}{16}$
2	0	0	0	$\frac{194}{16}$
3	0	0	0	$\frac{227}{16}$

Notamos aquí que la política óptima es única. Para explicar la política óptima ,vemos de la tabla que al comienzo del período 1, si el inventario tiene 0 unidades, hay que ordenar 3 unidades, en otro caso no ordenar; si el inventario en el período 2 es de 0 unidades, hay que ordenar 2 unidades, en otro caso no ordenar; y en el período 3 no ordenar. A manera de función, en la primer época de decisión, la política de decisión es

$$d_1^*(x) = \begin{cases} 0 \text{ si } x > 0 \\ 3 \text{ si } x = 0 \end{cases}.$$

El algoritmo anterior puede ser modificado para usar el índice de recompensa descontada esperada en lugar de la recompensa total esperada.. En ese caso, podemos considerar que trabajamos con nuevas funciones recompensa dadas por

$$r_t'(x_t, a_t) = \lambda^t r_t(x_t, a_t),$$

y obtenemos

$$w_N^{\pi}(x_N) = \lambda^N r_N(x_N),$$

y, para $t = N-1, N-2, \ldots 1$

$$w_t^*(x_t) = \max_{a \in A_{x_t}} \left\{ \lambda^t r_t(x_t, a) + \sum_{j \in X} w_{t+1}^*(j)\ p_t(j \mid x_t, a) \right\}.$$

Si ahora definimos $W_t^{\pi}(x_t) = \lambda^{-t} w_t^{\pi}(x_t)$, obtenemos

$$W_N^{\pi}(x_N) = r_N(x_N),$$

y, para $t = N-1, N-2, \ldots 1$

$$W_t^*(x_t) = \max_{a \in A_{x_t}} \left\{ r_t(x_t, a) + \lambda \sum_{j \in X} W_{t+1}^*(j)\ p_t(j \mid x_t, a) \right\}.$$

3.3.2 Optimalidad de políticas monótonas

En las secciones anteriores hemos hablado de condiciones bajo las cuales existe una política optima en el proceso de decisión Markoviano. En esta sección daremos condiciones adicionales que aseguran que las políticas óptimas son monótonas no decrecientes o no crecientes en los estados del sistema, y la importancia de su estudio consiste en que facilitan su implementación en el cálculo de los datos.

Una *política límite de control* es una política de Markov compuesta de reglas de decisión de la forma

$$d_t(x) = \begin{cases} a_1, & x < x^* \\ a_2, & x \geq x^*, \end{cases},$$

donde a_1 y a_2 son acciones distintas, y x^* es un límite de control. Tal política se interpreta como sigue: Cuando el estado del sistema es menor que x^*, es óptimo elegir la acción a_1, y cuando el estado del sistema es x^* o mayor, resulta óptimo elegir la acción a_2. La sencillez de su implementación radica en que para encontrar la política óptima, tenemos que determinar x^*.

Funciones superaditivas

Sean X y Y conjuntos parcialmente ordenados y $g(x, y)$ una función en $X \times Y$ con valores reales. Decimos que g es *superaditiva* si para $x^+ \geq x^-$ en X y $y^+ \geq y^-$ en Y,

$$g(x^+, y^+) + g(x^-, y^-) \geq g(x^+, y^-) + g(x^-, y^+). \tag{3.3.10}$$

Si se da la desigualdad contraria, decimos que g es *subaditiva*. Entonces $g(x,y)$ es subaditiva si $-g(x,y)$ es superaditiva. Algunas veces las funciones que satisfacen 3.3.10 se les llama también funciones supermodulares y la expresión 3.3.10 se conoce como la desigualdad cuadrangular.

Equivalentemente decimos que $g(x,y)$ es superaditiva si la función tiene diferencias monótonas crecientes; es decir,

$$g(x^+,y^+) - g(x^+,y^-) \geq g(x^-,y^+) - g(x^-,y^-). \tag{3.3.11}$$

Una consecuencia de 3.3.11 es que , cuando $X = Y = \mathbb{R}$, y $g(x,y)$ es doblemente diferenciable, entonces g es superaditiva siempre que

$$\frac{\delta^2 g(x,y)}{\delta x \delta y} \geq 0.$$

Desde nuestra perspectiva, el siguiente lema nos servirá para nuestro tema.

Lema 3.3.5 *Supongamos que g es una función superaditiva en $X \times Y$ y que para cada $x \in X$, $\max_{y \in Y} g(x,y)$ existe. Entonces*

$$f(x) = \max\left\{ y' \in \arg\max_{y \in Y} g(x,y) \right\}$$

es monótona no decreciente en x.

Demostración. Sea $x^+ \geq x^-$ y escogemos $y \leq f(x^-)$. Entonces por la definición de f,

$$g(x^-, f(x^-)) - g(x^-, y) \geq 0,$$

y por 3.3.10

$$g(x^-,y) + g(x^+, f(x^-)) \geq g(x^-, f(x^-)) + g(x^+,y).$$

Reescribiendo la segunda desigualdad como

$$g(x^+, f(x^-)) \geq g(x^+,y) + [g(x^-, f(x^-)) - g(x^-,y)]$$

y combinándola con la desigualdad anterior, tenemos que

$$g(x^+, f(x^-)) \geq g(x^+,y),$$

para toda $y \leq f(x^-)$. Consecuentemente, $f(x^+) \geq f(x^-)$ ■

Optimalidad de políticas monótonas

Si X representa el conjunto de los enteros no negativos y $A_x = A'$ para todo $x \in X$, definimos

$$q_t(k \mid x, a) = \sum_{j=k}^{\infty} p_t(j \mid x, a),$$

para $t = 1, ..., N-1$. La expresión anterior representa la probabilidad de que el estado en la época de decisión $t+1$ exceda el estado $k-1$.

Lema 3.3.6 *Sean* $\{x_j\}, \left\{x'_j\right\}$ *sucesiones valuadas en los reales y no negativas que satisfacen*

$$\sum_{j=k}^{\infty} x_j \geq \sum_{j=k}^{\infty} x'_j, \tag{3.3.12}$$

para toda k, *satisfaciendo la igualdad para* $k = 0$.

Supongamos $v_{j+1} \geq v_j$ *para* $j = 0, 1, ...,$ *entonces*

$$\sum_{j=0}^{\infty} v_j x_j \geq \sum_{j=0}^{\infty} v_j x'_j, \tag{3.3.13}$$

donde los límites en 3.3.13 existen y pueden ser infinito.

Demostración. Sea k arbitraria y $v_{-1} = 0$. Entonces

$$\begin{aligned}
\sum_{j=0}^{\infty} v_j x_j &= \sum_{j=0}^{\infty} x_j \sum_{i=0}^{j} (v_i - v_{i-1}) = \sum_{j=0}^{\infty} (v_j - v_{j-1}) \sum_{i=j}^{\infty} x_i \\
&= \sum_{j=1}^{\infty} (v_j - v_{j-1}) \sum_{i=j}^{\infty} x_i + v_0 \sum_{i=0}^{\infty} x_i \geq \sum_{j=1}^{\infty} (v_j - v_{j-1}) \sum_{i=j}^{\infty} x'_i + v_0 \sum_{i=0}^{\infty} x'_i \\
&= \sum_{j=0}^{\infty} v_j x'_j.
\end{aligned}$$

Como una consecuencia de esta última representación, ambos límites en 3.3.13 existen.

■

Notemos que el lema anterior nos permite derivar una propiedad importante de variables aleatorias ordenadas estocásticamente. Supongamos que Y y Y' denotan variables aleatorias con $x_j = P[Y = j]$ y $x'_j = P[Y' = j]$. Cuando 3.3.12 se cumple, Y es estocásticamente mayor que Y' (sección 3.1). Como consecuencia de 3.3.13, se sigue que para cualquier función no decreciente $f(j)$, $E[f(Y)] \geq E[f(Y')]$.

Proposición 3.3.7 *Supongamos que se alcanza el máximo en*

$$\max_{a \in A'} \left\{ r_t(x, a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x, a)\, w(j) \right\},$$

y que

1.- $r_t(x,a)$ *es no decreciente (no creciente) en* x *para todo* $a \in A'$ *y* $t = 1, ..., N-1$,

2.- $q_t(k \mid x,a)$ *es no decreciente en* x *para toda* $k \in X$, $a \in A'$ *y* $t = 1, ..., N-1$., *y*

3.- $r_N(x)$ *es no decreciente (no creciente) en* x.

Entonces $w_t^*(x)$ *es no decreciente (no creciente) en* x *para* $t = 1, ..., N$.

Demostración. Como $w_N^*(x) = r_N(x)$, el resultado se cumple para $t = N$ por 3.

Supongamos que $w_n^*(x)$ es no decreciente para $n = t+1, ..., N$. Por suposición existe una $a_x^* \in A'$ que alcanza el máximo en

$$w_t^*(x) = \max_{a \in A'} \left\{ r_t(x,a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x,a)\, w_{t+1}^*(j) \right\},$$

de tal manera que

$$w_t^*(x) = r_t(x,a_x^*) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x,a_x^*)\, w_{t+1}^*(j).$$

Sea $x' \geq x$. Por 1 y 2, la hipótesis de inducción, y el lema sobre suma de sucesiones aplicado con $x_j = p_t(j \mid x,a_x^*)$, $x_j' = p_t(j \mid x',a_x^*)$ y $v_j = w_{t+1}^*(j)$,

$$\begin{aligned} w_t^*(x) &\leq r_t(x',a_x^*) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x',a_x^*)\, w_{t+1}^*(j) \\ &\leq \max_{a \in A'} \left\{ r_t(x',a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x',a)\, w_{t+1}^*(j) \right\} = w_t^*(x'). \end{aligned}$$

Por lo que $w_t^*(x')$ es no decreciente, la hipótesis de inducción se satisface, y se sigue el resultado. ■

El siguiente teorema establece las condiciones bajo las cuales existen políticas óptimas monótonas. Notamos aquí que pueden existir otras políticas óptimas que no son monótonas.

Teorema 3.3.8 *Supongamos que para* $t = 1, ..., N-1$, *se cumple*

1.- $r_t(x,a)$ *es no decreciente en* x *para todo* $a \in A'$,

2.- $q_t(k \mid x,a)$ *es no decreciente en* x *para toda* $k \in X$, *y* $a \in A'$,

3.- $r_t(x,a)$ *es una función superaditiva (subaditiva) en* $X \times A'$,

4.- $q_t(k \mid x,a)$ *es una función superaditiva (subaditiva) en* $X \times A'$ *para toda* $k \in X$, *y*

5.- $r_N(x)$ *es no decreciente en* x.

Entonces existen reglas de decisión óptimas $d_t^*(x)$ *que son no decrecientes (no crecientes) en* x

para $t = 1, ..., N-1$.

Demostración. Hay que demostrar que

$$v_t(x,a) = r_t(x,a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x,a)\, w_t^*(j)$$

es superaditiva, siempre y cuando q y r lo sean. Por la condición 4, y la definición de superaditividad, para $x^- \leq x^+$ y toda $k \in X$,

$$\sum_{j=k}^{\infty} \left[p_t(j \mid x^-, a^-) + p_t(j \mid x^+, a^+)\right] \geq \sum_{j=k}^{\infty} \left[p_t(j \mid x^-, a^+) + p_t(j \mid x^+, a^-)\right].$$

Por la proposición anterior, $w_t(x)$ es no decreciente en x para toda t, por lo que aplicando el lema de suma de sucesiones tenemos,

$$\sum_{j=0}^{\infty} \left[p_t(j \mid x^-, a^-) + p_t(j \mid x^+, a^+)\right] w_t(j) \geq \sum_{j=0}^{\infty} \left[p_t(j \mid x^-, a^+) + p_t(j \mid x^+, a^-)\right] w_t(j).$$

Entonces para cada t, $\sum_{j=0}^{\infty} p_t(j \mid x,a)\, w_t(j)$ es superaditiva.

De la condición 3, $r_t(x,a)$ es superaditiva y, puesto que la suma de funciones superaditivas es superaditiva, $v_t(x,a)$ es superaditiva. El resultado se sigue del lema (0.11.5). El caso subaditivo se demuestra de manera similar. ■

Inducción monótona hacia atrás

Supongamos que existe una política óptima monótona para cada $t = 1, ..., N-1$, que X es finito y $A_x = A'$ para toda $x \in X$.

Algorithm 3.3.9 (de programación dinámica para el caso monótono.) *1.- Hacemos* $t = N$ *y*

$$w_N^*(x) = r_N(x) \text{ para toda } x \in X.$$

2.- Sustituimos $t-1$ *por* t, *fijamos* $x = 1$ *y* $A_1' = A'$

2a.- Hacemos

$$w_t^*(x) = \max_{a \in A_x'} \left\{ r_t(x,a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x,a)\, w_{t+1}^*(j) \right\}.$$

2b.- Hacemos

$$A_{x,t}^* = \arg\max_{a \in A_x'} \left\{ r_t(x,a) + \lambda \sum_{j=0}^{\infty} p_t(j \mid x,a)\, w_{t+1}^*(j) \right\}.$$

2c.- Si $x = M$, *ir a 3, de otra forma hacemos*

$$A_{x+1}' = \left\{ a \in A' : a \geq \max \left[a' \in A_{x,t}^* \right] \right\}.$$

2d.- Sustituimos $x+1$ *por* x*, y regresamos a 2a.*

3.- Si $t=1$ *paramos, de otra manera regresamos a 2.*

Este algoritmo difiere del algoritmo de programación dinámica en que la maximización es llevada a cabo sobre los conjuntos $A_x^{'}$ que se hacen cada vez más pequeños, al incrementar x, y por lo tanto reducimos el número de acciones que se necesitan evaluar en el paso 2. Si en algún x', A_x' contiene un solo elemento, entonces no se necesita hacer otra maximización puesto que esa acción será óptima para toda $x \geq x'$.En este caso, si solo queda una acción, digamos a^*, no necesitamos maximizar más en la iteración t y en lugar de esto, hacemos

$$w_t^*(x) = r_t(x, a^*) + \lambda \sum_{j \in X} p_t(j \mid x, a^*)\, w_{t+1}^*(j).$$

Por esta razón, tal algoritmo es particularmente utilizado en un modelo con dos acciones.

3.4 Problemas con horizonte infinito

3.4.1 Espacios de Banach y notación vectorial para los procesos de decisión Markovianos

Sea V el conjunto de las funciones reales acotadas en X, es decir, $v \in V$ si $v : X \to \mathbb{R}$ y existe una K_v tal que$\mid v(x) \mid \leq K_v$ para toda $x \in X$. Notemos que $\sup_{v \in V} K_v$ puede ser igual a $+\infty$. Para cada $v \in V$, definimos la norma del supremo como

$$\|v\| = \sup_{x \in X} \mid v(x) \mid .$$

Cuando X es finito, el supremo es alcanzado y nos referiremos a él como "max" en lugar de "sup". Como V es cerrado bajo la adición y la multiplicación por un escalar, y esta referido a una norma, se trata de un espacio lineal normado.

Decimos que un espacio lineal normado es completo si cada sucesión de Cauchy contiene un punto límite en ese espacio. El espacio lineal normado V junto con la norma es un espacio lineal normado completo o espacio de Banach. Cuando X es un subconjunto del espacio Euclidiano, V_M denota la familia de las funciones reales medibles acotadas en X. Para X finito o contable (relacionado con una topología discreta), todas las funciones valuadas en los reales son medibles de tal manera que V y V_M coinciden, pero cuando X es continuo, V_M es un subespacio propio de V. Sea $e \in V$ la función con todos los componentes igual a 1; esto es, $e(x) = 1\ \forall x \in X$.

Para X discreto, a menudo referimos los elementos de V como vectores y a los operadores lineales en V como matrices. Cuando usamos la norma en V, la norma matricial correspondiente esta dada por

$$\|H\| = \sup_{x \in X} \sum_{j \in X} |H(j \mid x)|.$$

donde $H(j \mid x)$ es la componente (x, j) - ésima de H.

Para X discreto, sea $|X|$ el número de elementos en X. Para $d \in D$, definimos $r_d(x)$ y $p_d(j \mid x)$ por

$$r_d(x) = r(x, d(x)) \text{ y } p_d(j \mid x) = p(j \mid x, d(x)).$$

Sea r_d el vector de dimensión $|X|$, con componente $x-$ésima $r_d(x)$ y P_d la matriz $|X| \times |X|$ con (x, j) ésima entrada dada por $p_d(j \mid x)$. Nos referimos a r_d como el vector de recompensas y P_d como la matriz de transición de probabilidad correspondiente a la regla de decisión Markoviana d. Para $0 \leq \lambda \leq 1$, $r_d + \lambda P_d v$ es igual a la recompensa total descontada esperada para un periodo con regla de decisión d y recompensa terminal v.

3.4.2 El algoritmo de programación dinámica hacia adelante

Una forma natural de trabajar un problema con horizonte infinito es aplicar un límite cuando el numero de épocas de decisión es N, de tal manera que

$$v_\lambda^\pi(x) = \lim_{N \to \infty} v_{\lambda,N}^\pi(x) = \lim_{N \to \infty} E_x^\pi \{ \sum_{t=1}^{N} \lambda^{t-1} r(x_t, a_t) \},$$

y para ello sería de gran utilidad que el algoritmo no estuviera formulado de atrás para adelante sino al revés, es decir, que cada término dependa del anterior y partiendo desde $t = 1$. Para esto haremos un cambio de variable. Antes de ello conviene hacer explícita una hipótesis que supondremos válida de aqui en adelante.

Hipótesis 3.4.1 *Existe una constante M tal que $0 \leq r(x, a) \leq M$, $\forall x \in X$ y $a \in A_x$*

Algorithm 3.4.2 *El algoritmo de programación dinámica hacia adelante es:*

a.-

$$\omega_1(x) = w_N(x) = \max_{\pi \in \Pi} r_N(x); \; x \in X. \tag{3.4.14}$$

b.- haciendo $w_t(x) = \omega_{N-t+1}(x)$ y $r_t(x, a) = r_{N-t+1}(x, a)$, con $t = 1, 2, ..., N$. Así tenemos:

$$\omega_{t+1}(x) = \max_{a \in A(x)} \{ r_{t+1}(x, a) + \lambda \sum_{y \in X} \omega_t(y) p(y \mid x, a) \}, \tag{3.4.15}$$

para $t = N-1, N-2, ..., 1$ y $\forall x \in X$.

A las funciones ω_t las llamaremos funciones de iteración de valores hacia adelante.

Lema 3.4.3 *Supongamos que X es discreta, $|r(x,a)| \leq M$ para toda $a \in A$, y $x \in X$, y $0 \leq \lambda \leq 1$. Entonces, para todo $v \in V$ y $d \in D$, $r_d + \lambda P_d v \in V$.*

Demostración. Como consecuencia de la suposición en $r(x,a)$, $\|r_d\| \leq M$ para toda $d \in D$, tal que $r_d \in V$. Cuando P_d es una matriz de probabilidad, $\|P_d\| = 1$, tal que $\|P_d v\| \leq \|P_d\|\|v\| = \|v\|$. Consecuentemente, $P_d v \in V$ para toda $v \in V$, y $r_d + \lambda P_d v \in V$. ■

Bajo la política $\pi = (d_1, d_2, ...) \in \prod$, la componente (x,j) –ésima de la matriz de transición de probabilidad del paso t P_π^t satisface

$$P_\pi^t(j \mid x) = \left[P_{d_t} P_{d_{t-1}} \cdots P_{d_1}\right](j \mid x) = P^\pi(X_{t+1} = j \mid X_1 = x).$$

La esperanza con respecto a la cadena de Markov correspondiente a esta política se calcula conforme a

$$E_x^\pi\{v(X_t)\} = P_\pi^{t-1} v(x) = \sum_{j \in X} P_\pi^{t-1}(j \mid x)\, v(j),$$

para $v \in V$ y $1 \leq t < \infty$.

Como consecuencia de esta representación de la esperanza, y de la definición de v_λ^π, para $0 \leq \lambda \leq 1$,

$$v_\lambda^\pi = \sum_{t=1}^{\infty} \lambda^{t-1} P_\pi^{t-1} r_{d_t},$$

asumiendo la existencia del límite.

3.4.3 Ecuación de optimalidad

Definición 3.4.4 *Decimos que una función $v \in V$, donde V representa al espacio de funciones de Banach (espacio lineal normado completo), es una solución de la ecuación de optimalidad λ descontada si:*

$$v(x) = \max_{a \in A(x)} \{r(x,a) + \lambda \sum_{y \in X} v(y) p(y \mid x, a)\, ; \forall x \in X.$$

Definición 3.4.5 *Para cada $v \in V$, definimos al operador*

$$Tv(x) = \max_{a \in A(x)} \{r(x,a) + \lambda \sum_{y \in X} v(y) p(y \mid x, a)\}; \forall x \in X. \tag{3.4.16}$$

y para $d \in D$,

$$T_d v(x) = r(x,d) + \lambda \sum_{y \in X} v(y) p(y \mid x, d)\, ; \forall x \in X.$$

Hacemos notar aquí que estamos trabajando con el conjunto de políticas estacionarias, i.e., $\pi = d^{\infty} = (d, d, ...)$. También tenemos que $T^t v = T(T^{t-1}v)$, y similarmente para T_d. Además es fácil mostrar que $T^t v \in V$ y $T_d^t v \in V$ para $v \in V$, $t = 1, 2, ...$ y el algoritmo formado por 3.4.14 y 3.4.15 se puede expresar como

$$\omega_1 = \omega, \tag{3.4.17}$$

$$\omega_t = T^t \omega; \forall x \in X. \tag{3.4.18}$$

Observación 3.4.6 *En términos del operador T, la ecuación de optimalidad queda expresada como $v = Tv$, $v \in V$. El hecho de que A_x, $x \in X$, sea finito nos garantiza la existencia de $d \in D$ tal que para $v \in V$,*

$$Tv(x) = r(x, d) + \lambda \sum_{y \in X} v(y) p(y \mid x, d); \forall x \in X,$$

es decir, $Tv(x) = T_d v(x)$.

Lema 3.4.7 *Supóngase que la Hipótesis 3.4.1 se cumple. Si $v \in V$, y satisface $v \leq Tv$, entonces*

$$v \leq v_{\lambda}^{\pi}(x) = \lim_{N \to \infty} E_x^{\pi}\{\sum_{t=1}^{N} \lambda^{t-1} r_t(X_t, Y_t)\} = \lim_{N \to \infty} v_{N,\lambda}^{\pi}(x); \ \forall \pi \in \Pi.$$

De aquí, $v \leq v_{\lambda}^{*}(x)$.

Demostración. Primero vemos que la relación $v \leq Tv$ implica que

$$v \leq r(x, a) + \lambda \sum_{y \in X} v(y) p(y \mid x, a); \forall x \in X.$$

Ahora sean x y π elementos arbitrarios de X y Π respectivamente. Como v es una función de x_{t+1} entonces

$$E_x^{\pi}(v(x_{t+1} \mid h_t)) = \sum_{y \in X} v(y) p(y \mid x_t, d_t(h_t)),$$

por lo que

$$\begin{aligned} E_x^{\pi}\{\lambda^{t+1} v(x_{t+1}) \mid h_t, a_t\} &= \lambda^{t+1} \sum_{y \in X} v(y) p(y \mid x_t, d_t(h_t) \\ &= \lambda^t \{r(x_t, a_t) + \lambda \sum_{y \in X} v(y) p(y \mid x_t, d_t(h_t)\} - \lambda^t r(x_t, a_t) \\ &\geq \lambda^t v(x_t) - \lambda^t r(x_t, a_t). \end{aligned}$$

La última desigualdad es debido a 3.4.17. Como $v(x_t) = E_x^\pi\{v(x_t) \mid h_t, a_t\}$;

$$\lambda^t r(x_t, a_t) \geq E_x^\pi\{\lambda^t v(x_t) - \lambda^{t+1} v(x_{t+1}) \mid h_t, a_t\};$$

tomando esperanza de ambos lados de la desigualdad obtenemos

$$E_x^\pi\{\lambda^t r(x_t, a_t)\} \geq E_x^\pi\{\lambda^t v(x_t) - \lambda^{t+1} v(x_{t+1}) \mid h_t, a_t\}.$$

Ahora, sumando sobre $t = 1, 2, ..., n$;

$$E_x^\pi \sum_{t=1}^{n} \lambda^t r(x_t, a_t) \geq v(x) - \lambda^{n+1} E_x^\pi\{v(x_{n+1})\};$$

haciendo $n \to \infty$ y por el hecho de que $v \in V$ y π es arbitraria, obtenemos que $v_\lambda^\pi(x) \geq v(x); \forall \pi \in \Pi, x \in X$.En particular $v_\lambda^*(x) \geq v(x)$. ■

Lema 3.4.8 *Suponemos la Hipótesis 3.4.1, entonces T y T_d, con $d \in D$, son operadores de contracción sobre V con módulo λ,i.e., para cualquier par de funciones $u, v \in V$,*

(a) $\|Tu - Tv\| \leq \lambda \|u - v\|$.

(b) $\|T_d u - T_d v\| \leq \lambda \|u - v\|$,

donde $\|v\| = \sup_{x \in X} \mid v(x) \mid$.

Demostración. Para $x \in X$, $a \in A(x)$, $v, v \in V$, tenemos

$$\begin{aligned}
r(x,a) + \lambda \sum_{y \in X} v(y) p(y \mid x, a) &= r(x,a) + \lambda \sum_{y \in X} v(y) p(y \mid x, a) + \lambda \sum_{y \in X} u(y) p(y \mid x, a) \\
-\lambda \sum_{y \in X} u(y) p(y \mid x, a) &= r(x,a) + \lambda \sum_{y \in X} u(y) p(y \mid x, a) \\
+\lambda \sum_{y \in X} [v(y) - u(y)] p(y \mid x, a) &\leq r(x,a) + \lambda \sum_{y \in X} u(y) p(y \mid x, a) + \lambda \sup_{y \in X} \mid v(y) - u(y) \mid .
\end{aligned}$$

Tomando máximo sobre A_x en ambos lados de la desigualdad y de 3.4.16 tenemos

$$Tv(x) - Tu(x) \leq \lambda \sup_{x \in X} \mid v(x) - u(x) \mid = \lambda \|v - u\|; \forall x \in X.$$

De manera completamente análoga obtenemos

$$Tu(x) - Tv(x) \leq \lambda \|u - v\|; \forall x \in X.$$

De las dos desigualdades llegamos a que

$$\mid Tv(x) - Tu(x) \mid \leq \lambda \|v - u\|; \forall x \in X.$$

Tomando supremo de la parte izquierda de la desigualdad terminamos la demostración de (a). De manera similar se demuestra la parte (b) ■

Lema 3.4.9 *Bajo la Hipótesis 3.4.1, para cada $d \in D$, $v_\lambda^{d^\infty}(x)$ es el único punto fijo del operador T_d en V.*

Demostración. Sea $d \in D$. Usando la notación $r(x, d_t) = r_{d_t}$, P_{d_t} denota la matriz de transición usando la regla de decisión d_t, y $P_\pi^0 = I$, tenemos

$$v_\lambda^\pi(x) = E_x^\pi\{\sum_{t=1}^{\infty}\lambda^{t-1}r(x_t, a_t)\} = \sum_{t=1}^{\infty}\lambda^{t-1}P_\pi^{t-1}r_{d_t} = r_{d_1}+\lambda P_{d_1}(r_{d_2}+\lambda P_{d_2}r_{d_3}+\lambda P_{d_2}P_{d_3}r_{d_4}+.....),$$

de tal manera que

$$v_\lambda^\pi(x) = r_{d_1} + \lambda P_{d_1}v_\lambda^{\pi'}(x),$$

donde $\pi' = (d_2, d_3, ...)$.

Esta ecuación nos muestra que la recompensa total esperada descontada correspondiente a la política π es igual a la recompensa descontada en un período en que utiliza la política d_1 y recibe la recompensa total esperada descontada de la política π' como recompensa terminal. Si se expresa como notación de componente y no como matriz, tendríamos

$$v_\lambda^\pi(x) = r_{d_1(x)} + \sum_{y\in X}\lambda v_\lambda^{\pi'}(y)p(y \mid x, d_1),$$

y si π es estacionaria, i.e., $\pi = d^\infty = (d, d, ...)$, tenemos que $\pi = \pi'$ y la ecuación quedaría como

$$v_\lambda^{d^\infty}(x) = r_{d(x)} + \sum_{y\in X}\lambda v_\lambda^{d^\infty}(y)p(y \mid x, d) = r_d + \lambda P_d v_\lambda^{d^\infty} = T_d v_\lambda^{d^\infty}(x);\ \forall x \in X.$$

De aquí, $v_\lambda^{d^\infty}(x)$ es punto fijo de T_d. Para probar la unicidad comenzamos escribiendo

$$v = r_d + \lambda P_d v$$

como

$$(I - \lambda P_d)\, v = r_d.$$

Debido a que $\|P_d\| = 1$ y $\lambda = \|\lambda P_d\| \geq \sigma(\lambda P_d)$ (ver apéndice). Entonces, por el apéndice, para $0 \leq \lambda < 1$, $(I - \lambda P_d)^{-1}$ existe, de tal manera que

$$v = (I - \lambda P_d)^{-1} r_d = \sum_{1}^{\infty}\lambda^{t-1}P_d^{t-1}r_d = v_\lambda^{d^\infty}.$$

■

Lema 3.4.10 *Supongamos que* $0 \leq \lambda < 1$ *y* $u, v \in V$. *Entonces para cualquier* $d \in D$,
a.Si $u \geq 0$, *entonces* $(I - \lambda P_d)^{-1} u \geq 0$ *y* $(I - \lambda P_d)^{-1} u \geq u$.
b.Si $u \geq v$, *entonces* $(I - \lambda P_d)^{-1} u \geq (I - \lambda P_d)^{-1} v$; *y*
c.Si $u \geq 0$, *entonces* $u^T (I - \lambda P_d)^{-1} \geq 0$ *y* $u^T (I - \lambda P_d)^{-1} \geq u^T$.

Demostración. Puesto que $\sigma(\lambda P_d) < 1$ (ver apéndice), y la no negatividad de todos los elementos de P_d se observa que

$$(I - \lambda P_d)^{-1} u = u + \lambda P_d u + \lambda^2 P_d u + \cdots \geq u \geq 0.$$

La parte b) se sigue de reemplazar u por $u - v$ en a), y la parte c) se sigue de a) por transposición. ■

Teorema 3.4.11 *Supongamos se cumple la Hipótesis 3.4.1 Entonces*

(a) $\omega_t \to v_\lambda^*$ *donde* $\omega_0 \in V$ *es arbitraria y* ω_t *con* $t = 1, 2, \ldots$ *son las funciones de iteración de valores definidas por el algoritmo hacia adelante.*

(b) v_λ^* *es la única solución en* V *de la ecuación de optimalidad, i.e., cuando llegas al óptimo se convierte en punto fijo de* T.

$$v_\lambda^*(x) = \max_{a \in A(x)} \{r(x,a) + \lambda \sum v_\lambda^*(y) p(y \mid x, a)\};\ x \in X. \tag{3.4.19}$$

De manera equivalente $v_\lambda^* = T v_\lambda^*$.

Demostración. Por el lema (0.12.8 (a)), el operador T es de contracción, de esta manera utilizando el teorema de punto fijo de Banach (ver apéndice), tenemos que existe un único $v^* \in V$ tal que

$$Tv^* = v^*, \tag{3.4.20}$$

es decir, existe una única solución de la ecuación de optimalidad. Además tenemos que la sucesión $\{\omega_t\}$ definidas en el algoritmo de programación dinámica hacia adelante convergen a v^* cuando $t \to \infty$. Por lo tanto para concluir la prueba del teorema, sólo basta probar que $v^* = v_\lambda^{d^*}$.

Por el lema (0.12.7), 3.4.20 implica que $v^* \leq v_\lambda^{d^*}$. Para probar la desigualdad contraria, sea $d^* \in D$ una política estacionaria tal que $v^* = Tv^*$ (ver observación (0.12.6)). Así por el lema (0.12.9) $v^*(.) = v_\lambda^{d^*}(.)$. Por último, como $v^*(.) \leq v_\lambda^\pi(.), \forall \pi \in \Pi$, tenemos que $v^*(.) \geq v_\lambda^{d^*}(.)$. Por lo tanto $v^* = v_\lambda^{d^*}$. ■

Observación 3.4.12 *Usando el lema (0.12.9) y las propiedades del operador* T_d, *es fácil mostrar que para cada* $d \in D$ *y* $v \in V$,*se cumple* $v_\lambda^d(x) = \lim_{t \to \infty} T_d^t v(x),\ x \in X$.

Como lo muestra el siguiente teorema, el hecho de que v_λ^* sea solución de la ecuación de optimalidad nos permite caracterizar a las políticas óptimas. Partiendo de la existencia de una política estacionaria $d \in D$ tal que $v_\lambda^* = T_d v_\lambda^*$, tenemos:

Teorema 3.4.13 *Supóngase que la Hipótesis 3.4.1 se cumple.*

(a) Una política estacionaria $d^* \in D$ es óptima si y solo si maximiza el lado derecho de 3.4.19, i.e.,

$$v_\lambda^*(x) = r(x, d^*) + \lambda \sum v_\lambda^*(y) p(y \mid x, d^*)\}; \; x \in X. \tag{3.4.21}$$

En general,

(b) Una política $\pi \in \Pi$ es óptima si y sólo si $v_\lambda^\pi(x)$, $x \in X$, satisface la ecuación de optimalidad.

Demostración.) Sea $d^* \in D$ una política que satisface 3.4.21.

En la última parte de la demostración del Teorema (0.12.11) se mostró que $v_\lambda^*(x) = v_\lambda^{d^*}(x), x \in X$; lo cual implica que d^* es óptima.

Ahora, sea $d^* \in D$ una política óptima, es decir, $v_\lambda^*(x) = v_\lambda^{d^*}(x), x \in X$. De aquí y del lema (0.12.9), $v_\lambda^*(x) = T_{d^*} v_\lambda^*(x)$, lo cual prueba la parte (a) del teorema.

(b) Sea $\pi \in \Pi$ una política óptima, i.e., $v_\lambda^\pi(x) = v_\lambda^*(x)$, $x \in X$. Por el teorema (0.12.11 (b)), $Tv_\lambda^\pi(x) = v_\lambda^\pi(x)$, $x \in X$.

Para demostrar la implicación inversa solo basta probar que $v_\lambda^\pi(x) \leq v_\lambda^*(x)$, $x \in X$. Para esto supongamos que $v_\lambda^\pi(x)$ satisface la ecuación de optimalidad, es decir, $v_\lambda^\pi(x) = Tv_\lambda^\pi(x)$, $x \in X$. En particular, $v_\lambda^\pi(x) \leq Tv_\lambda^\pi(x)$, y por el lema (0.12.7), $v_\lambda^\pi(x) \leq v_\lambda^*(x), x \in X$, lo cual implica que π es óptima. ■

Definición 3.4.14 *Para $v \in V$, decimos que una regla de decisión $d \in D$ es $v-$mejorada si*

$$d_v \in \arg\max_{d \in D} \{r_d + \lambda P_d v\},$$

Equivalentemente,

$$r_{d_v} + \lambda P_{d_v} v = \max_{d \in D} \{r_d + \lambda P_d v\} \;\; o \; T_{d_v} v = Tv,$$

que significa lo mismo.

En notación de componentes, d_v es $v-$mejorada si para toda $x \in X$;

$$r(x, d_v(x)) + \sum_{j \in X} \lambda p(j \mid x, d_v(x)) v(x) = \max_{a \in A_x} \left\{ r(x, a) + \sum_{j \in X} \lambda p(j \mid x, a) v(j) \right\}.$$

Aunque se le llama $v-$mejorada, en realidad lo que quiere decir es que la recompensa descontada esperada de d_v es por lo menos tan grande como v, pero no necesariamente mayor que v. El mejoramiento estricto ocurre cuando d_v es $v-$mejorada y

$$r_{d_v}(x') + \lambda P_{d_v} v(x') > v(x'),$$

para al menos una x', para la cual $P_{d_v}(x, x') > 0$ para alguna $x \in X$.

Las reglas de decisión $d \in D$, que son v_λ^* mejoradas las llamaremos conservadoras de tal manera que,

$$T_{d^*} v_\lambda^* \equiv r_{d^*} + \lambda P_{d^*} v_\lambda^* = v_\lambda^*, \tag{3.4.22}$$

o alternativamente si,

$$d^* \in \arg\max_{d \in D} \{r_d + \lambda P_d v_\lambda^*\}.$$

El siguiente teorema provee un método para identificar políticas estacionarias óptimas.

Teorema 3.4.15 *Sea S discreto, y supongamos que el supremo es alcanzado en $r_d + \lambda P_d v$, para toda $v \in V$. Entonces*

a. Existe una regla de decisión conservadora $d^* \in D$

b. Si d^* es conservadora, la política estacionaria determinista $(d^*)^\infty$ es óptima; y

c. $v_\lambda^* = \sup_{d \in D} v_\lambda^\infty$

Demostración. La parte a) se sigue notando que $v_\lambda^* \in V$ y que el supremo en $r_d + \lambda P_d v$ es alcanzado. Por el teorema (0.12.11(b)), v_λ^* es la única solución de $Tv = v$. Por lo que, de 3.4.22,

$$v_\lambda^* = T v_\lambda^* = r_{d^*} + \lambda P_{d^*} v_\lambda^* = T_{d^*} v_\lambda^*,$$

por lo cual, del lema (0.12.9),

$$v_\lambda^{(d^*)^\infty} = v_\lambda^*.$$

La parte c) es una consecuencia inmediata de la parte b). ■

3.4.4 Iteración de valores

Aunque existen varios métodos como son el método de iteración de políticas y programación lineal, con sus respectivas variantes cada uno, el método más adecuado para estos problemas es el de iteración de valores o de aproximaciones sucesivas.

Como Hipótesis adicional a la anterior, supondremos al espacio de estados finito. El método se refiere a la implementación del algoritmo 3.4.17 y 3.4.18, del cual podemos decir que consiste en iterar sucesivamente el operador T definido por 3.4.16 y cuya convergencia esta demostrada en el Teorema (0.12.11 (a)).

Tasas de convergencia

Sea $\{y_n\} \subset V$ una sucesión que converge a y^*, es decir, $\lim_{n\to\infty} \|y_n - y^*\| = 0$. Decimos que $\{y_n\}$ converge en orden (por lo menos) $\alpha, \alpha > 0$ si existe una constante $K > 0$ para la cual

$$\|y_{n+1} - y^*\| \leq K\|y_n - y^*\|^\alpha, \tag{3.4.23}$$

para $n = 1, 2, ...$ La convergencia lineal corresponde a α por lo menos 1; la cuadrática a α por lo menos dos y decimos que $\{y_n\}$ converge superlinealmente si

$$\limsup_{n\to\infty} \frac{\|y_{n+1} - y^*\|}{\|y_n - y^*\|} = 0.$$

Si $\{y_n\}$ converge a orden α, definimos la tasa de convergencia $\hat{K}$ como la más pequeña K tal que 3.4.23 se satisface para toda n, y notamos que mientras $\hat{K}$ es más pequeña la convergencia es más rápida. El requerimiento de que 3.4.23 se cumpla para toda n puede causar que esta medida sea insensitiva a las verdaderas propiedades de convergencia de una sucesión. Para adquirir una mejor visión de la velocidad de convergencia, se define la tasa de convergencia promedio asintótica (TCPA) como

$$\limsup_{n\to\infty} \left[\frac{\|y_n - y^*\|}{\|y_0 - y^*\|}\right]^{\frac{1}{n}}.$$

Notamos que esta definición puede extenderse a series que convergen a orden mayor que 1 elevando la cantidad en el denominador a α^n.

Dada una función $f(n)$ no negativa en los reales, definida en los enteros, decimos que la sucesión $\{y_n\}$ es $O(f(n))$ cuando

$$\limsup_{n\to\infty} \frac{\|y_n - y^*\|}{f(n)}$$

sea finito. En este caso se escribe $y_n = y^* + O(f(n))$. Cuando $f(n) = \beta^n$, con $0 < \beta < 1$ decimos que la convergencia es geométrica a tasa β.

Supongamos que una sucesión converge linealmente con TCPA igual a Φ. Esto significa que para cualquier $\varepsilon > 0$, existe una N tal que, para $n \geq N$,

$$\|y_n - y^*\| \leq (\Phi + \varepsilon)^n \|y_0 - y^*\|,$$

de tal manera que la convergencia es $O((\Phi + \varepsilon)^n)$.

Ahora lo extendemos a algoritmos. En este libro, representamos algoritmos iterativos por mapeos $T : V \to V$. Dado un valor inicial y_0, el algoritmo genera iteraciones $Ty_0, T^2y_0, ...$ En general, $y_n = T^n y_0$. Distinguimos tasas y ordenes de convergencia globales y locales . Decimos que un algoritmo converge localmente con un orden específico, tasa de convergencia, o TCPA si la sucesión $\{T^n y_0\}$ para y_0 fija converge con

ese orden, tasa o TCPA. Decimos que el algoritmo converge globalmente a un orden específico, tasa o TCPA si converge localmente para toda y_0. Entonces, las tasas y ordenes de convergencia algorítmicos globales miden el desempeño del peor caso de los valores del algoritmo, mientras los ordenes y tasas de convergencia local miden el desempeño de un valor inicial particular.

Interpretamos la tasa promedio asintótica de convergencia de un algoritmo como sigue. Supongamos que deseamos conocer el número n de iteraciones, requeridos para reducir el error $\|y_n - y^*\|$ por una fracción ϕ del error inicial $\|y_0 - y^*\|$. Si la tasa promedio asintótica de convergencia iguala ρ, encontramos n resolviendo $\phi^{\frac{1}{n}} = \rho$ o $n \approx \frac{\log(\phi)}{\log(\rho)}$ iteraciones. Puesto que ϕ y ρ son menores que 1, mientras ρ este más cerca de 1, más iteraciones se requieren para obtener una reducción dada en el error. Por ejemplo, si $\rho = 0.9$, requiere 22 iteraciones para reducir el error por un factor de 10 ($\phi = 0.1$).

Cuando comparamos dos métodos iterativos, con TCPA ρ_1 y ρ_2, el número de iteraciones requeridas para reducir el error inicial por un monto fijo usando el método 1, n_1, esta relacionado con el segundo método, n_2, por

$$\frac{n_1}{n_2} = \frac{\log(\rho_2)}{\log(\rho_1)}.$$

Regresando al tema del algoritmo de iteración de valores, primero veremos cual es la tasa de convergencia del algoritmo. Si comenzamos con $\omega_1 = 0$,

$$\|\omega_1 - v_\lambda^*\| \leq \frac{\lambda^n M}{1-\lambda}.$$

Esto es fácil de ver notando primero que por el lema (0.12.8), el operador T^n, satisface la relación

$$\|T^n\omega - T^n v\| \leq \lambda^n \|\omega - v\|; \quad v, \omega \in V.$$

Tomando el caso particular donde $v = v_\lambda^*$ y $\omega = 0$, y el teorema (0.12.11(b)) ,

$$\|\omega_n - v_\lambda^*\| \leq \lambda^n \|v_\lambda^*\| \leq \frac{\lambda^n M}{1-\lambda}$$

Esto muestra que la tasa de convergencia del algoritmo es geométrica.

Definición 3.4.16 *Sea $\varepsilon > 0$, decimos que una política π_ε^* es $\varepsilon-$óptima, si*

$$v_\lambda^{\pi_\varepsilon^*}(x) \leq v_\lambda^*(x) + \varepsilon, \forall x \in X.$$

Teorema 3.4.17 *Supongamos que el espacio de estados X es finito o numerable, entonces para toda $\varepsilon > 0$ existe una política estacionaria determinista $\varepsilon-$óptima.*

Demostración. Del teorema (0.12.11), $Tv_\lambda^* = v_\lambda^*$. Escojamos $\varepsilon > 0$ y seleccionamos $d_\varepsilon \in D$ que satisface

$$r_{d_\varepsilon} + \lambda P_{d_\varepsilon} v_\lambda^* \geq \sup_{d \in D} \{r_d + \lambda P_d v_\lambda^*\} - (1-\lambda)\varepsilon e = v_\lambda^* - (1-\lambda)\varepsilon e.$$

donde $e \in V$ es el vector en el cual todos sus componentes son iguales a 1.

Puesto que $v_\lambda^{(d_\varepsilon)} = (I - \lambda P_{d_\varepsilon})^{-1} r_{d_\varepsilon}$, reagrupando términos y multiplicando por $(I - \lambda P_{d_\varepsilon})^{-1}$ se llega a

$$v_\lambda^{(d_\varepsilon)} \geq v_\lambda^* - \varepsilon e,$$

de tal manera que d_ε es óptima. ■

Antes de ver una variante del algoritmo de iteración de valores, haremos un breve recordatorio a las diferentes tasas de con vergencia. A esta variante se le conoce como el algoritmo de iteración de valores Gauss Seidel y su tasa de convergencia es mayor que la del algoritmo de iteración de valores.

Algoritmo de Iteración de Valores

Paso 1: Hacemos $t = 1$. Seleccionamos una función $\omega_1 \in V$, y especificamos $\varepsilon > 0$.

Paso 2: Calculamos ω_{t+1} mediante

$$\omega_{t+1}(x) = \max_{a \in A(x)} \{r(x,a) + \lambda \sum_{y \in X} \omega_t(y) p(y \mid x, a)\}, \ \forall x \in X.$$

Paso 3: Si

$$\|\omega_{t+1} - \omega_t\| < \frac{\varepsilon(1-\lambda)}{2\lambda} \tag{3.4.24}$$

continuamos con el Paso 4, de lo contrario regresamos al Paso 2 incrementando t en 1.

Paso 4: Elegimos $d_\varepsilon \in D$ tal que $\forall x \in X$:

$$\omega_{t+1} = r(x, d_\varepsilon) + \lambda \sum_{y \in X} \omega_t(y) p(y \mid x, d_\varepsilon) = \max_{a \in A(x)} \left\{ r(x,a) + \lambda \sum_{y \in X} \omega_t(y) p(y \mid x, a) \right\}, \tag{3.4.25}$$

o de manera equivalente

$$T_{d_\varepsilon} \omega_t = \omega_{t+1} \tag{3.4.26}$$

Observación 3.4.18 *Del teorema (0.12.11(a)) se sigue que existe un $L < \infty$ tal que 3.4.24 se cumple para todo $t \geq L$.*

Teorema 3.4.19 *Sea $\varepsilon > 0$, $\omega_1 \in V$ arbitrario y fijo y ω_n, $n \geq 1$ las funciones de iteración de valores. Entonces,*

(a) 3.4.24 implica que $\|\omega_{t+1} - v_\lambda^*\| < \frac{\varepsilon}{2}$;

(b) La política estacionaria definida por 3.4.25 es $\varepsilon-$óptima.

Demostración. (a) Supóngase que 3.4.24 se cumple para algún t. Por el teorema (0.12.11(b)), la relación 3.4.19 y Lema (0.12.8(a)) tenemos que

$$\begin{aligned}\|v_\lambda^* - \omega_{t+1}\| &= \|Tv_\lambda^* - \omega_{t+1}\| \leq \|Tv_\lambda^* - T\omega_{t+1}\| + \|T\omega_{t+1} - \omega_{t+1}\| \\ &= \|Tv_\lambda^* - T\omega_{t+1}\| + \|T\omega_{t+1} - T\omega_t\| \leq \lambda\|v_\lambda^* - \omega_{t+1}\| + \lambda\|\omega_{t+1} - \omega_t\|.\end{aligned}$$

Por lo tanto,

$$\|v_\lambda^* - \omega_{t+1}\| \leq \frac{\lambda}{1-\lambda}\|\omega_{t+1} - \omega_t\|.$$

De aquí y de la relación 3.4.24 obtenemos

$$\|\omega_{t+1} - v_\lambda^*\| < \frac{\varepsilon}{2}. \tag{3.4.27}$$

(b) Nuevamente, supongamos que se cumple 3.4.24 para alguna t, y sea $d_\varepsilon \in D$ una política que satisface 3.4.25. Siguiendo argumentos similares a la prueba de (a) con $v_\lambda^{d\varepsilon}(.) = v_\lambda(d_\varepsilon, .)$ en lugar de v_λ^*, y usando 3.4.26, el Lema (0.12.9) y Lema (0.12.8) obtenemos

$$\|v_\lambda^{d_\varepsilon} - \omega_{t+1}\| < \frac{\varepsilon}{2}. \tag{3.4.28}$$

Por último, observamos que

$$\|v_\lambda^{d_\varepsilon} - v_\lambda^*\| \leq \|v_\lambda^{d_\varepsilon} - \omega_{t+1}\| + \|\omega_{t+1} + v_\lambda^*\| < \varepsilon,$$

donde la última desigualdad se sigue de 3.4.27 y 3.4.28. Esto implica que d_ε es una política óptima. ■

Algoritmo de iteración de valores Gauss-Seidel

1. Especificamos $\omega^0(x)$ para toda $x \in X, \varepsilon > 0$, y fijamos $n = 0$.

2. Fijamos $j = 1$ e ir a 2(a).

   2(a). Calculamos $\omega^{n+1}(x_j)$ por

$$\omega^{n+1}(x_j) = \max_{a \in A_{x_j}} \left\{ r(x_j, a) + \lambda \left[\sum_{i<j} p(x_i \mid x_j, a)\, \omega^{n+1}(x_i) + \sum_{i \geq j} p(x_i \mid x_j, a)\, \omega^n(x_i) \right] \right\}. \tag{3.4.29}$$

   2(b). Si $j = N$, ir al paso 3. De otra manera incrementar j en 1 e ir a 2(a).

3. Si

$$\|\omega^{n+1} - \omega^n\| < \frac{\varepsilon(1-\lambda)}{2\lambda}$$

   ir al paso 4. Si no, incrementar j en 1 y regresar al paso 2.

4. Para cada $x \in X$, escogemos

$$d^{\varepsilon}(x) \in \arg\max_{a \in A_x} \left\{ r(x,a) + \sum_{j \in X} \lambda p(j \mid x,a)\, \omega^{n+1}(j) \right\}$$

y nos detenemos.

Representamos ahora una modificación a través del paso 2 como una partición regular de matrices. Sea d la regla de decisión correspondiente a las acciones maximizadoras, obtenidas cuando evaluamos 3.4.29 para $x_1, x_2, ..., x_N$, y descomponemos P_d como $P_d = P_d^L + P_d^U$, donde

$$P_d^L = \begin{pmatrix} 0 & 0 & \cdot & \cdot & \cdot & 0 \\ p_{21} & 0 & & & & 0 \\ p_{31} & p_{32} & 0 & & & \\ \cdot & & & \cdot & & \\ \cdot & & & & \cdot & \\ p_{N1} & & & p_{N,N-1} & 0 & \end{pmatrix}, P_d^U = \begin{pmatrix} p_{11} & p_{12} & \cdot & & \cdot & p_{1N} \\ 0 & p_{22} & \cdot & & \cdot & p_{2N} \\ 0 & & p_{33} & & & p_{3N} \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 & p_{NN} \end{pmatrix}.$$

Se sigue que el paso 2 puede escribirse como

$$\omega^{n+1} = \left(I - \lambda P_d^L\right)^{-1} \left(\lambda P_d^U\right) \omega^n + \left(I - \lambda P_d^L\right)^{-1} r_d.$$

Haciendo $Q_d = \left(I - \lambda P_d^L\right)$ y $R_d = \lambda P_d^U$, es fácil ver que (Q_d, R_d) es una partición regular de $I - \lambda P_d$. Por lo que el valor iterado Gauss Seidel puede expresarse como

$$\omega^{n+1} = Q_d^{-1} R_d \omega^n + Q_d^{-1} r_d.$$

Notemos que a pesar de representar el algoritmo Gauss Seidel en términos de la inversa de Q_d, no hay necesidad de evaluar explícitamente esta inversa, simplemente representa los cálculos en el paso 2.

Ahora daremos un teorema general de convergencia para métodos basados en particiones regulares.

Teorema 3.4.20 *Supongamos que (Q_d, R_d) es una partición regular de $I - \lambda P_d$ para toda $d \in D$ y que*

$$\alpha \equiv \sup_{d \in D} \| Q_d^{-1} R_d \| < 1, \tag{3.4.30}$$

entonces:

a. Para toda $\omega^0 \in V$, el esquema iterativo

$$\omega^{n+1} = \max_{d \in D} \left\{ Q_d^{-1} r_d + Q_d^{-1} R_d \omega^n \right\} \equiv T\omega^n \tag{3.4.31}$$

converge a v_λ^*.

b. La cantidad v_λ^* es el único punto fijo de T.

c. La secuencia $\{\omega^n\}$ definida por 3.4.31 converge globalmente con orden uno a una tasa

menor o igual que α, su tasa promedio global asintótica de convergencia es menor o igual que α, y

converge globalmente $O(\beta^n)$ donde $\beta \leq \alpha$.

Demostración. Primero demostraremos que 3.4.30 implica que T es un mapeo de contracción. Asumamos que para una $x \in X$ y $u, v \in V$, que $Tv(x) - Tu(x) \geq 0$. Sea

$$d_v \in \arg\max_{d \in D} \left\{ Q_d^{-1} r_d + Q_d^{-1} R_d v \right\},$$

entonces,

$$0 \leq Q_{d_v}^{-1} r_{d_v}(x) + Q_{d_v}^{-1} R_{d_v} v(x) - Q_{d_v}^{-1} r_{d_v}(x) - Q_{d_v}^{-1} R_{d_v} u(x) \leq \|Q_{d_v}^{-1} R_{d_v}\| \|v - u\|.$$

Aplicando el mismo argumento en el caso $0 \leq Tu(x) - Tv(x)$, y usando 3.4.30 de nuevo, se establece que

$$\|Tv - Tu\| \leq \alpha \|v - u\|, \tag{3.4.32}$$

de tal manera que T es un mapeo de contracción en V. Y por el teorema del punto fijo de Banach (ver apéndice), $\{\omega^n\}$ converge al único punto fijo de T que denotamos por v^*.

Ahora mostraremos que $v^* = v_\lambda^*$. Puesto que v^* es un punto fijo de T, para toda $d \in D$

$$v^* \geq Q_d^{-1} r_d(x) + Q_d^{-1} R_d v^*;$$

consecuentemente,

$$\left(I - Q_d^{-1} R_d\right) v^* \geq Q_d^{-1} r_d.$$

De 3.4.30, $\sigma\left(Q_d^{-1} R_d\right) < 1$, de tal manera que por el corolario del apéndice, $\left(I - Q_d^{-1} R_d\right)^{-1}$ y satisface

$$\left(I - Q_d^{-1} R_d\right)^{-1} = \sum_0^\infty \left(Q_d^{-1} R_d\right)^n.$$

De la definición de Q_d y R_d, $\left(I - Q_d^{-1} R_d\right)^{-1} \geq 0$. Por lo tanto,

$$v^* \geq \left(I - Q_d^{-1} R_d\right)^{-1} Q_d^{-1} r_d = (Q_d - R_d)^{-1} r_d = (I - \lambda P_d)^{-1} r_d = v_\lambda^{d^\infty}.$$

Así que, por el teorema (0.12.9), existe una política estacionaria óptima, $v^* \geq v_\lambda^*$, pero por suposición existe una $d^* \in D$ tal que $v^* = v_\lambda^{(d^*)^\infty}$, por lo cual $v^* = v_\lambda^*$.

Ahora bien, por a), para cualquier $\omega^0 \in V$ las iteraciones de valores Gauss-Seidel satisfacen

$$\|\omega^{n+1} - v_\lambda^*\| = \|T\omega^n - Tv_\lambda^*\| \leq \|Q_{d_v}^{-1} R_{d_v}\| \|\omega^n - v_\lambda^*\| \leq \alpha \|\omega^n - v_\lambda^*\|, \qquad (3.4.33)$$

donde α se define como 3.4.30. Supongamos que existe $\hat{d} \in D$ tal que $\|Q_{\hat{d}}^{-1} R_{\hat{d}}\|$ es el valor más pequeño que cumple la igualdad en 3.4.33 para toda n, entonces no existe $d \in D$ tal que $\|Q_d^{-1} R_d\| < \|Q_{\hat{d}}^{-1} R_{\hat{d}}\| \leq \alpha$, pero esta es una contradicción por ser α el supremo, por lo que la tasa de convergencia es menor o igual a α.

Iterando 3.4.33 desde $n = 0$, dividiendo ambos lados por $\|\omega^0 - v_\lambda^*\|$ y tomando la raíz n–ésima mostramos que

$$\limsup \left[\frac{\|\omega^n - v_\lambda^*\|}{\|\omega^0 - v_\lambda^*\|} \right]^{\frac{1}{n}} \leq \alpha,$$

y por la misma razón de que α es un supremo se tiene que la tasa de convergencia promedio asintótica (TCPA) es menor o igual que α.

Por último, iteramos 3.4.33 desde $n = 0$, y dividimos ambos lados por $\beta^n = \|Q_{d_v}^{-1} R_{d_v}\|^n$ para mostrar que

$$\limsup \frac{\|\omega^n - v_\lambda^*\|}{\beta^n} \leq \|\omega^0 - v_\lambda^*\|,$$

por lo que la iteración converge globalmente $O(\beta^n)$, donde $\beta \leq \alpha$. ■

Proposición 3.4.21 *Sea P una matriz de probabilidad de transición y (Q_1, R_1) y (Q_2, R_2) particiones regulares de $I - \lambda P$, donde $0 \leq \lambda < 1$. Entonces, si $R_2 \leq R_1 \leq \lambda P$,*

$$\|Q_2^{-1} R_2\| \leq \|Q_1^{-1} R_1\|.$$

Demostración. Para $i = 1$ o 2, $I - Q_i = \lambda P - R_i$ tal que $\lambda P \geq R_i \geq 0$ implica que $\lambda P \geq I - Q_i \geq 0$. Consecuentemente, $1 > \lambda = \|\lambda P\| \geq \|I - Q_i\|$. Por el corolario del apéndice se deduce que $Q_i^{-1} = I + (I - Q_i) + (I - Q_i)^2 + \cdots$. Debido a que $R_2 \leq R_1$, $I - Q_2 \geq I - Q_1$. Combinando estas dos observaciones se implica que $Q_2^{-1} \geq Q_1^{-1}$. Puesto que $(I - \lambda P) e \geq 0$,

$$Q_2^{-1} (I - \lambda P) e \geq Q_1^{-1} (Q_1 - R_1) e \geq 0,$$

e implica que $0 \leq Q_2^{-1} R_2 e \leq Q_1^{-1} R_1 e$. Por lo cual $\|Q_2^{-1} R_2 e\| \leq \|Q_1^{-1} R_1 e\|$. Como $Q_i^{-1} R_i \geq 0$, para $i = 1, 2$,

$$\|Q_i^{-1} R_i e\| = \|Q_i^{-1} R_i\|,$$

de donde se sigue el resultado. ■

Teorema 3.4.22 *Para una $\omega^0 \in V$ arbitraria, las iteraciones del la iteración de valores Gauss Seidel $\{\omega_{GS}^n\}$ convergen a v_λ^*. Más, la convergencia es global de orden uno, a una tasa menor o igual que λ; su tasa promedio asintótica global es menor o igual que λ, y converge globalmente $O(\beta^n)$ con $\beta \leq \lambda$.*

Demostración. Aplicando la proposición anterior, con $R_1 = \lambda P_d$ y $R_2 = \lambda P_d^U$, implica que, para toda $d \in D$,

$$\| \left(I - \lambda P_d^L\right)^{-1} \lambda P_d^U \| \leq \|\lambda P_d\| = \lambda < 1.$$

Consecuentemente, por el teorema (0.12.20), $\{\omega_{GS}^n\}$ converge a v_λ^* a las tasas supuestas.
■

3.4.5 Apéndice

3.4.6 Espacios lineales

Decimos que Q es una transformación lineal en el espacio lineal normado V si para cualquiera escalares α, β, y u,v elementos en V, $Q(\alpha u + \beta v) = \alpha Qu + \beta Qv$. Cuando X es discreto, nos referimos a Q como una matriz.

Definición 3.4.23 *Una transformación lineal Q en V es acotada, si existe una constante $K > 0$ tal que, para todo $v \in V$,*

$$\|Qv\| \leq K\|v\|.$$

Proposición 3.4.24 *Sea V un espacio de Banach y $Q \in L(V)$, entonces $L(V)$ (definido como el conjunto de transformaciones lineales acotadas en V), cuya norma es*

$$\|Q\| = \sup\{\|Qv\| : \|v\| \leq 1, v \in V,\},$$

es un espacio de Banach.

Definición 3.4.25 *Se define al radio espectral de Q, denotado por $\sigma(Q)$, por*

$$\sigma(Q) = \lim_{n\to\infty} \|Q^n\|^{\frac{1}{n}}.$$

Cuando $Q \in L(V)$, este límite existe. Claramente, $\sigma(Q) \geq 0$. En general, si $P, Q \in L(V)$,

$$\|PQ\| \leq \|P\|\|Q\|,$$

de tal manera que $\|Q^n\|^{\frac{1}{n}} \leq \|Q\|$, por lo que $\sigma(Q) \leq \|Q\|$, y se puede ver que

$$\sigma(PQ) \leq \sigma(P)\sigma(Q).$$

Teorema 3.4.26 *Sea Q una transformación lineal acotada en un espacio de Banach V, y supongamos que $\sigma(I-Q)<1$. Entonces Q^{-1} existe y satisface*

$$Q^{-1}=\lim_{N\to\infty}\sum_{0}^{N}(I-Q)^n.$$

Demostración. Hipótesis, existe una $b<1$ tal que $\sigma(I-Q)<b<1$. Como consecuencia de la definición de $\sigma(I-Q)$, dado $\varepsilon>0$, existe una N^* tal que, para $n\geq N^*$,

$$\|(I-Q)^n\|^{\frac{1}{n}}<b+\varepsilon<1,$$

tal que

$$\|(I-Q)^n\|<(b+\varepsilon)^n. \tag{3.4.34}$$

Ahora sea $U_N=\sum_{0}^{N}(I-Q)^n$. Entonces para $N>M\geq N^*$,

$$\|U_N-U_M\|=\|\sum_{M+1}^{N}(I-Q)^n\|\leq\sum_{M+1}^{N}\|(I-Q)^n\|\leq\sum_{M+1}^{N}(b+\varepsilon)^n.$$

Por lo tanto $\{U_N\}$ es una sucesión de Cauchy. Como $L(V)$ es un espacio de Banach, entonces existe una U^* en $L(V)$ que satisface

$$\lim_{N\to\infty}\|U_N-U^*\|=0.$$

Ahora demostraremos que este límite es igual a Q^{-1}. Como

$$\|I-QU_N\|=\|I-\left[I-(I-Q)^{N+1}\right]\|\leq\|(I-Q)^{N+1}\|,$$

se sigue de 3.4.34 que

$$\|I-QU^*\|=\lim_{N\to\infty}\|I-QU_N\|=0.$$

Similarmente, $\|I-U^*Q\|=0$, por lo que concluimos que $U=Q^{-1}$. ■

Puesto que $\|I-Q\|>\sigma(I-Q)$, el siguiente resultado es una consecuencia inmediata

Corolario 3.4.27 *Sea Q una transformación lineal acotada en el espacio de Banach V, y supongamos que $\sigma(Q)<1$. Entonces $(I-Q)^{-1}$ existe y satisface*

$$(I-Q)^{-1}=\lim_{N\to\infty}\sum_{0}^{N}Q^n.$$

Teorema 3.4.28 *(Teorema del punto fijo de Banach) Supongamos que U es un espacio de Banach y que $T : U \to U$ es una contracción. Entonces*

a. Existe una única v^* en U tal que $Tv^* = v^*$; y

b. Para una arbitraria ω^0 en U, la sucesión $\{\omega^n\}$ definida por

$$\omega^{n+1} = T\omega^n = T^{n+1}\omega^0, \tag{3.4.35}$$

converge a v^*.

Demostración. *inamos a $\{\omega^n\}$ por 3.4.35. Entonces para cualquier $m \geq 1$,*

$$\|\omega^{n+m} - \omega^n\| \leq \sum_0^{m-1} \|\omega^{n+k+1} - \omega^{n+k}\| = \sum_0^{m-1} \|T^{n+k}\omega^1 - T^{n+k}\omega^0\| \leq \sum_0^{m-1} T^{n+k}\|\omega^1 - \omega^0\|$$

$$= \frac{\lambda^n(1-\lambda^m)}{(1-\lambda)}\|\omega^1 - \omega^0\|. \tag{3.4.36}$$

Puesto que $0 \leq \lambda < 1$, se sigue de 3.4.36 que $\{\omega^n\}$ es una sucesión de Cauchy; esto es, para n suficientemente grande, $\|\omega^{n+m} - \omega^n\|$ puede hacerse arbitrariamente pequeña. De la completez de U, se sigue que $\{\omega^n\}$ tiene límite $v^* \in U$.

Ahora mostraremos que v^* es un punto fijo de T. Usando las propiedades de normas y de contracción, se sigue que

$$0 \leq \|Tv^* - v^*\| \leq \|Tv^* - \omega^n\| + \|\omega^n - v^*\| = \|Tv^* - T\omega^{n-1}\| + \|\omega^n - v^*\|$$

$$\leq \lambda\|v^* - \omega^{n-1}\| + \|\omega^n - v^*\|.$$

Puesto que el $\lim_{n\to\infty} \|\omega^n - v^*\| = 0$, ambas cantidades en la parte derecha de la desigualdad anterior pueden hacerse arbitrariamente pequeñas escogiendo la n suficientemente grande. Consecuentemente, $\|Tv^* - v^*\| = 0$, de donde concluimos que $Tv^* = v^*$.

Sea $\{\omega^n\}$ una sucesión definida como 3.4.35. Entonces, para $m \geq 1$,

$$\begin{aligned} \|\omega^{n+m} - \omega^n\| &\leq \sum_0^{m-1} \|\omega^{n+k+1} - \omega^{n+k}\| = \sum_0^{m-1} \|T^{n+k}\omega^1 - T^{n+k}\omega^0\| \\ &\leq \sum_0^{m-1} \lambda^{n+k}\|\omega^1 - \omega^0\| = \frac{\lambda^n(1-\lambda^m)}{(1-\lambda)}\|\omega^1 - \omega^0\|. \end{aligned}$$

Como $0 \leq \lambda \leq 1$, se sigue de 3.4.36 que $\{\omega^n\}$ es una sucesión de Cauchy; esto es, para n lo suficientemente grande, $\|\omega^{n+m} - \omega^n\|$ puede hacerse lo suficientemente pequeño. De la completez de U, se sigue que

$$\begin{aligned} 0 &\leq \|Tv^* - v^*\| \leq \|Tv^* - \omega^n\| + \|\omega^n - v^*\| = \|Tv^* - T\omega^{n-1}\| + \|\omega^n - v^*\| \\ &\leq \lambda\|v^* - \omega^{n-1}\| + \|\omega^n - v^*\|. \end{aligned}$$

Debido a que $\lim_{n\to\infty} \|\omega^n - v^*\| = 0$, *ambas cantidades en la parte derecha de la desigualdad de arriba pueden hacerse lo suficientemente pequeñas escogiendo una* n *lo suficientemente grande. Consecuentemente,* $\|Tv^* - v^*\| = 0$, *de donde concluimos que* $Tv^* = v^*$. *Como* $\{v^n\}$ *es una sucesión de Cauchy, la sucesión tiene un límite y por lo tanto es único.* ■

Capítulo 4

Política óptima de reclamos en un seguro multiperíodo

4.1 Introducción

En este capítulo analizaremos el problema, que presentamos en el capítulo 1, relativo a las condiciones que permiten optimizar la utilidad de un asegurado que ha contratado un seguro de daños multiperíodo, ya que a diferencia de los contratos a un solo período, la prima del asegurado aumenta si el asegurado incurre en un reclamo. Los resultados que presentaremos a continuación, fueron expuestos en el artículo optimal multi-period insurance contracts y abarcan varios aspectos. En la sección 4.3 se supone que el asegurado está sujeto a un contrato dado, establecido exógenamente, y se deduce que la forma de una política óptima de reclamos para maximizar la utilidad total del asegurado, a largo plazo, consiste en determinar un nivel crítico y^* (mayor que el deducible) y hacer la reclamación sólo en caso de que el daño exceda ese nivel crítico.

La sección 4.4 está dedicada a analizar qué tipo de contrato es el mejor para un asegurado que ha decidido realizar una política de reclamos como la que se describió en el párrafo anterior. Se demuestra que el contrato óptimo es de cobertura total por encima de un deducible positivo.

Empezaremos reestableciendo el modelo secuencial que corresponde al problema.

4.2 El modelo

Es usual clasificar a los compradores de seguros de automóviles en un número finito de categorías de riesgo: $1, 2, \ldots, J$. Para esta clasificación, se toman en cuenta diversos aspectos como la edad, la experiencia de conductores, el estatus económico y la historia de reclamos que ha realizado el asegurado en el pasado, entre otros. Aquí vamos a

suponer que todos los factores distintos a la historia de reclamos están dados y son fijos.

El sistema que se aplica más frecuentemente para considerar dicha historia en la determinación de la categoría de riesgo, por el método conocido como "bonus-malus", que consiste en que si el asegurado está ubicado en la categoría j y realiza algún reclamo durante el período t, entonces para el período $t+1$ será colocado en la categoría $j+1$. Si en cambio no realiza ningún reclamo, es movido a la categoría $j-1$. Si el asegurado pertenece a la categoría $j=1$, permanecerá ahí aun cuando no haga reclamos. En el otro extremo, si pertenece a la categoría más alta J, al realizar reclamos saldrá del contrato. Esto lo contemplamos en el modelo matemático suponiendo que existe un estado $J+1$ en el que el asegurado prefiere quedarse sin seguro por el alto nivel de las primas o, de manera alternativa, la aseguradora le niega el seguro.

La categoría j determina el precio de las primas que cobrará la compañía. Llamaremos $I_j(x)$ a la función que indica la indemnización que la aseguradora paga por un daño de x pesos cuando el asegurado está ubicado en la categoría j. Evidentemente, I_j es una función no-decreciente del daño y satisface las condiciones:

$$0 \leq I_j(x) \leq x, \text{ y}$$

$$I_j(0) = 0.$$

Para facilitar la exposición, asumiremos que $I_j(x)$ es continua y diferenciable excepto, posiblemente, en un número finito de puntos. Esta indemnización depende de la cobertura que ofrece el contrato y por tanto identificaremos el tipo de contrato con la función de indemnización correspondiente. El precio de un contrato I_j, es decir, la prima correspondiente, será denotada por $r_j(I_j)$. Como a mayor j, mayor riesgo, supondremos que $r_j(\cdot) > r_{j-1}(\cdot)$.

El comprador tiene dos tipos de decisiones que tomar: (1) elegir qué contrato comprar; y (2) dado el contrato que ha comprado, determinar una política de reclamos. El segundo tipo de decisión sólo tiene sentido en un seguro multiperíodo y es el que discutiremos a continuación.

4.3 Política óptima de reclamos dado un contrato I_j

4.3.1 Suposiciones iniciales y notación.

Sea X_j la variable aleatoria que indica el monto del daño ocurrido durante un período para un asegurado de la categoría j, y supondremos que su distribución de probabilidad es independiente del tiempo t. Sea C_t el ingreso del asegurado en el período t y $u(\cdot)$ su función de utilidad.

Para que las recompensas sean finitas basta suponer que $E\left[u\left(C_t - X_J\right)\right]$ es finita (Como vimos en el capítulo 2, la manera en que el asegurado mide el valor de un proyecto económico aleatorio es mediante el valor esperado de su función de utilidad, por lo que la utilidad pasa a ser la recompensa del asegurado). Como $E\left[u\left(C_t - X_J\right)\right]$ es finita, las $E[u(C - X_j)]$ son finitas para toda $j \leq J.$, debido a que las x_j son no negativas y que $F_j(x) \geq F_{j+1}(x)$, donde F_j es la distribución de X_j.

Asumimos que todos los daños son reparados, por lo que el consumo del asegurado en el período τ, c_τ, es $[C_t - r_j(I_{j_\tau}) - x_j + I_{j_\tau}(x_j)]$ si el asegurado sufre un daño x_j y realiza el reclamo, y $[C_t - r_j(I_{j_\tau}) - x_j]$ si el asegurado sufre el daño y no reclama.

Sea $V_{jt}(y)$ la máxima utilidad esperada descontada del asegurado del período t en adelante, condicionado a que en t ha tenido un daño y, y que él siempre sigue una estrategia de reclamos óptima.

Supongamos que conocemos la recompensa esperada descontada óptima de $t+1$ en adelante, cuando el asegurado se encuentra en la categoría j, denotada por $W_{j,t+1}$. Por el algoritmo de programación dinámica, la recompensa esperada descontada óptima de t en adelante suponiendo que ocurrió un daño de monto y, esta dada por

$$W_{j_t}(y) = \max\{u(C_t - r_j - y) + \beta\left[W_{j-1,t+1}\left(X_{j-1}\right)\right], u(C_t - r_j - y + I_{j_t}(y)) + \beta\left[W_{j+1,t+1}\left(X_{j+1}\right)\right]\},$$

donde β es el factor de descuento, $0 < \beta < 1$.

Para facilitar la discusión siguiente, tomamos

$$V_{j_t}(y) = \max\{u(C_t - r_j - y) + \beta W_{j-1,t+1}, u(C_t - r_j - y + I_{j_t}(y)) + \beta W_{j+1,t+1}\},$$

de manera que $W_{j_t}\left(y\right) = E[V_{j_t}(Y_j)]$. Es decir, $V_{j_t}\left(y\right)$ representa la máxima utilidad del asegurado de t en adelante suponiendo que en ese período sufrió un daño y y que siempre ha aplicado una política óptima. Para determinar una política óptima, se deben evaluar las W_{j_t}. Hay dos alternativas para lograrlo.

1) Suponer un horizonte de planeación T finito, pero arbitrariamente grande. En este caso $W_{j,T+1} = 0$ para toda j y las W_{j_t} se pueden calcular usando programación dinámica.

2) Suponer que el horizonte de planeación es infinito tomando $C_t = C, \forall t$, y que las primas y las indemnizaciones no varían con el tiempo, es decir, un modelo estacionario.

Aquí desarrollaremos la segunda vía.

4.3.2 La política óptima

Para determinar la política óptima tenemos que evaluar W_{j_t}. Para ello asumimos que todas las C_t son iguales a C y que las primas y las indemnizaciones no varían en los distintos períodos. Entonces tenemos

$$V_{j_t}(y) = \max\{u(C - r_j - y) + \beta W_{j-1}, u(C - r_j - y + I_j(y)) + \beta W_{j+1}\}. \qquad (4.3.1)$$

Así que, de acuerdo a 4.3.1, se realiza un reclamo si

$$h_{j_t}(y) \equiv u(C - r_j - y) - u(C - r_j - y + I_j(y)) \leq \beta(W_{j+1} + W_{j-1}). \qquad (4.3.2)$$

El siguiente Lema cumple condiciones suficientes para asegurar que existe un valor y_j en el que $h(y_j) = 0$.

Lema 4.3.1 *La función $h(y)$ definida por*

$$h(y) = u(C - r_j - y) - u(C - r_j - y + I_j(y)),$$

es continua y no-creciente.

Demostración. La continuidad y diferenciabilidad (excepto, posiblemente, en un número finito de puntos) se sigue de las suposiciones hechas sobre u e $I_j(y)$. Derivando h obtenemos:

$$\begin{aligned} h'(y) &= -u'(C - r_j - y) + u'(C - r_j - y + I_j(y))(1 - I_j'(y)) \\ &= -u'(C - r_j - y) + u'(C - r_j - y + I_j(y)) - I_j'(y)u'(C - r_j - y + I_j(y)). \end{aligned}$$

Debido a que u' es decreciente y a que $I_j(y) \geq 0$,

$$-u'(C - r_j - y) + u'(C - r_j - y + I_j(y)) \leq 0.$$

También la no negatividad de I' y de u' implica que

$$-I_j'(y)u'(C - r_j - y + I_j(y)) \leq 0.$$

La no negatividad de I' se deduce de que I es una función continua y no decreciente respecto al daño y. Así $h_j'(y) \leq 0$ y por lo tanto $h(y)$ es no creciente ■

Como $h_j(y)$ representa la diferencia de la utilidad del asegurado por cobrar o no cobrar la suma asegurada $I_j(y)$, y como es no creciente y continua, entonces existe un valor del daño que llamaremos y_j en que esta diferencia valdrá cero, por lo que para cada categoría j, el reclamo será realizado sólo si el daño y es mayor o igual a y_j, es decir que solo realizaremos el cobro si $y \geq y_j$. El siguiente Teorema nos permite asegurar que podemos obtener las W_j usando el método de iteración de valores.

Teorema 4.3.2 *Las funciones $W_j, j = 1, ..., J$, son finitas y únicas y pueden calcularse por el método de iteración de valores.*

Demostración. Sea Ω el conjunto $\{1, ..., J\}$, y sea $C(\Omega)$ el conjunto de todas las funciones $\phi : \Omega \to \mathbb{R}$. Sea $W = (W_1, ..., W_J) \in C = \times_1^J C(\Omega)$. Consideremos la transformación $T_j(W)$ definida por

$$T_j(W) = E[\max\{u(C - r_j - y) + \beta W_{j-1}, u(C - r_j - y + I_j(y)) + \beta W_{j+1}\}], j = 1, ...; J.$$

Ahora, basta mostrar que la transformación $T_j(W)$ es una contracción, entonces por 3.4.19 se sigue que existen funciones únicas $W_j^*, j = 1, ..., J$, que satisfacen

$$W_j^* = E\left[\max\{u(C - r_j - y) + \beta W_{j-1}^*, u(C - r_j - y + I_j(y)) + \beta W_{j+1}^*\}\right], j = 1, ...; J.$$

Para demostrar que T_j es una contracción, necesitamos demostrar que para cualesquiera $W^1, W^2 \in C$

$$\max_j \mid T_j(W^1) - T_j(W^2) \mid \leq \beta \max_i \mid W_i^1 - W_i^2 \mid,$$

para ello, notamos que

$$\begin{aligned} & T_j(W^1) - T_j(W^2) \\ &= E\left[\max\{u(C - r_j - y) + \beta W_{j-1}^1, u(C - r_j - y + I_j(y)) + \beta W_{j+1}^1\}\right] \\ &- E\left[\max\{u(C - r_j - y) + \beta W_{j-1}^2, u(C - r_j - y + I_j(y)) + \beta W_{j+1}^2\}\right]. \end{aligned} \tag{4.3.3}$$

Como vimos del Lema anterior, $u(C - r_j - y) - u(C - r_j - y + I_j(y))$ es continua y no creciente, de donde se sigue que 4.3.3 se puede escribir como

$$\begin{aligned} T_j(W^1) - T_j(W^2) &= \int_0^{y_j^1} [u(C - r_j - y) + \beta W_{j-1}^1] dF_j(y) \\ &+ \int_{y_j^1}^{\infty} [u(C - r_j - y + I_j(y)) + \beta W_{j+1}^1] dF_j(y) \\ &- \int_0^{y_j^2} [u(C - r_j - y) + \beta W_{j-1}^2] dF_j(y) \\ &- \int_{y_j^2}^{\infty} [u(C - r_j - y + I_j(y)) + \beta W_{j+1}^2] dF_j(y), \end{aligned} \tag{4.3.4}$$

donde y_j^1 y y_j^2 son los valores críticos correspondientes a $T_j(W^1)$ y $T_j(W^2)$, respectivamente. Como el valor crítico y_j^2 es tal que maximiza la segunda esperanza en el lado derecho de 4.3.3, se sigue que si sustituimos y_j^2 por y_j^1 en 4.3.4, entonces esa esperanza deja de ser un máximo y la diferencia entre las integrales aumentarán (un poco), y por lo tanto

$$\begin{aligned} T_j(W^1) - T_j(W^2) &\leq \int_0^{y_j^1} [u(C - r_j - y) + \beta W_{j-1}^1] dF_j(y) \\ &+ \int_{y_j^1}^{\infty} [u(C - r_j - y + I_j(y)) + \beta W_{j+1}^1] dF_j(y) \\ &- \int_0^{y_j^1} [u(C - r_j - y) + \beta W_{j-1}^2] dF_j(y) \\ &- \int_{y_j^1}^{\infty} [u(C - r_j - y + I_j(y)) + \beta W_{j+1}^2] dF_j(y) \\ &= \beta\left[F_j(y_j^1)(W_{j-1}^1 - W_{j+1}^2) + (1 - F_j(y_j^1))(W_{j-1}^1 - W_{j+1}^2)\right], \end{aligned} \tag{4.3.5}$$

donde $F_j(y) = \int_0^y dF_j(y)$.

De la misma manera

$$T_j(W^2) - T_j(W^1)$$

$$\begin{aligned}
&= E\left[\max\{u(C-r_j-y)+\beta W_{j-1}^2, u(C-r_j-y+I_j(y))+\beta W_{j+1}^2\}\right] \quad (4.3.6)\\
&\quad -E\left[\max\{u(C-r_j-y)+\beta W_{j-1}^1, u(C-r_j-y+I_j(y))+\beta W_{j+1}^1\}\right]\\
&\leq \int_0^{y_j^2}[u(C-r_j-y)+\beta W_{j-1}^2]dF_j(y)\\
&\quad +\int_{y_j^2}^{\infty}[u(C-r_j-y+I_j(y))+\beta W_{j+1}^2]dF_j(y)\\
&\quad -\int_0^{y_j^2}[u(C-r_j-y)+\beta W_{j-1}^1]dF_j(y)\\
&\quad -\int_{y_j^2}^{\infty}[u(C-r_j-y+I_j(y))+\beta W_{j+1}^1]dF_j(y)\\
&= \beta\left[F_j(y_j^2)(W_{j-1}^2-W_{j+1}^1)+(1-F_j(y_j^2))(W_{j-1}^2-W_{j+1}^1)\right].
\end{aligned}$$

De 4.3.5 y 4.3.6 se sigue que

$$|T_j(W^1) - T_j(W^2)| \leq \beta |W_{j-1}^1 - W_{j+1}^2|, \forall j = 1, ..., J,$$

y en particular

$$\max_j |T_j(W^1) - T_j(W^2)| \leq \beta \max_i |W_i^1 - W_i^2|,$$

lo que completa la demostración. ■

Con base en las W_j, la estrategia de reclamos óptima puede evaluarse a partir de 4.3.2 . De esta relación vemos que las y_j deben satisfacer

$$u(C-r_j-y_j) = u(C-r_j-y_j+I_j(y_j)) - \beta(W_{j-1}-W_{j+1}), \quad j=1,...,J; \quad (4.3.7)$$

donde $W_0 = W_1$ y $W_{J+1} = (1-\beta)^{-1}E[u(C-X_J)]$.

Nota. El valor de W_{J+1} se obtiene si asumimos que una vez que el asegurado ha alcanzado la categoría $J+1$, ya no sale de ahí, o dicho de otro modo, a $J+1$ lo consideramos un estado absorbente. En este caso, en cada período, su utilidad esperada es $E[u(C-X_J)]$ y por lo tanto su utilidad esperada descontada en horizonte infinito es $(1-\beta)^{-1}E[u(C-X_J)]$. Alternativamente podemos asumir que el comprador permanece en la categoría $J+1$ solo un período y siempre se mueve a la categoría J el siguiente período (no puede hacer el reclamo si no tiene seguro). En este caso $W_{J+1} = E[u(C-X_J)] + \beta W_J$. En ambas alternativas se obtiene el mismo resultado.

Además, como el bienestar de un asegurado en la categoría $j-1$ es mayor que el de un asegurado en la categoría $j+1$ debido a que las primas que el asegurado de la

categoría $j-1$ tiene que pagar son más bajas; entonces $W_{j-1} \geq W_{j+1}\ \forall j = 1, ..., J$. El caso $W_{j-1} = W_{j+1}$ para algún j es trivial puesto que implica que el sistema "Bonus-malus" no es efectivo para el conductor de la categoría j, así que asumiremos la desigualdad estricta. Entonces se sigue de 4.3.7, de $I_j(\cdot) \geq 0, I_j(0) = 0$, y de que u es creciente, que los valores críticos y_j son estrictamente positivos; es decir, para que se cumpla 4.3.7 , $u(C - r_j - y_j + I_j(y_j)) - u(C - r_j - y_j) > 0$, y para esto último, y_j debe ser estrictamente positivo. Como $W_j = E[V_j(Y)]$ y debido a que para $Y < y_j$ el asegurado no realiza ningún reclamo y para $Y \geq y_j$ si realiza el reclamo, se sigue que

$$W_j = \int_0^{y_j} [u(C - r_j - y) + \beta W_{j-1}] dF_j(y)$$

$$+ \int_{y_j}^{\infty} [u(C - r_j - y + I_j(y)) + \beta W_{j+1}] dF_j(y),\ j = 1, ..., J. \tag{4.3.8}$$

Así que las W_j, para $j = 1, ..., J$ representan la recompensa descontada esperada óptima para un asegurado que inicie en la categoría j.

Ahora analizaremos el problema de determinar el contrato óptimo. Para ello, en la siguiente sección eliminamos la suposición de que el contrato esta exógenamente determinado y estudiamos algunas propiedades de los contratos de seguros.

4.4 (b) Contratos de seguros óptimos

Vamos a modificar un resultado dado por Arrow (visto al final del capítulo 2) que asegura que el contrato de indemnización óptima para un período, debe ser de cobertura total por encima de un deducible. Queremos analizar la conveniencia de este tipo de contratos para el modelo multiperíodo y ver que también otorga esa misma cobertura total por encima de un deducible, pero solo cuando es por daños sobre un valor crítico y lo abreviaremos como CTEDSVC. Formalmente, lo anterior quiere decir que el contrato es igual a $I_j(x) = x - m_j$, para $x > y_j$; donde y_j es el valor crítico del daño y m_j es el deducible. La forma del contrato es arbitraria para $x \leq y_j$.

Se asume que el asegurado puede comprar cualquier contrato a un precio proporcional a su valor actuarial, es decir

$$r_j = (1 + \lambda_j) \int_{y_j}^{\infty} I_j(x) dF_j(x), \tag{4.4.9}$$

donde $\lambda_j > 0$ representa la tasa de interés utilizada para traer a valor presente el valor actuarial del contrato y se asume que λ_j satisface $\lambda_j \leq \lambda_{j+1}$. Esta desigualdad es necesaria puesto que en otro caso puede pasar que un asegurado de la categoría $j+1$ pague una prima más baja que un asegurado de la categoría j.

En lo que sigue mostraremos porque el contrato óptimo $I_j(\cdot)$ es de la forma CTEDSVC. La idea de la prueba es similar a la de Arrow (1963). Arrow mostró que en el modelo de un período, es óptimo tener $(I(x) - x)$ constante para todo daño x si $I(x) > 0$. Revisemos la demostración de Arrow para un solo período.

Supongamos que $I_j(\cdot)$ es un contrato óptimo. Consideremos una perturbación en $I_j(\cdot)$, es decir, $\bar{I}_j(\cdot) = I_j(\cdot) + g(\cdot)$, donde $g(\cdot)$ y su derivada $g'(\cdot)$ son pequeñas y $g(x) = 0$ para $0 \leq x \leq y_j$. Denotando por $L_j(g)$ el efecto de la perturbación g en la utilidad esperada del asegurado y asumiendo que $I_j(\cdot) + g(\cdot)$ es también un contrato factible, entonces

$$L_j(g) \equiv \int_0^\infty u(C - x + I_j(x) + g(x) - r_j)dF_j(x) - \int_0^\infty u(C - x + I_j(x) - r_j)dF_j(x) \tag{4.4.10}$$

deberá ser no positivo si el valor actuarial de g es cero (es decir si $\int_0^\infty g(x)dF_j(x) = 0$). Esto lo vemos porque si no fuera así, la utilidad esperada con el contrato $I_j(\cdot)$ sería mayor. Si llamamos $x_0 = C - x + I_j(x) - r_j$, y como u es diferenciable en todo x, entonces el incremento de u por un aumento Δx, se puede representar como

$$\Delta u = u'(x_0)\Delta x + o(\Delta x),$$

por lo que si $\Delta x = g(x)$, entonces

$$\Delta u = u(x_0 + g(x)) - u(x_0) = u'(x_0) * g(x) + o(g(x)).$$

Sustituyendo lo anterior en $L_j(g)$,tenemos

$$L_j(g) = \int_0^\infty g(x)u'(C - x + I_j(x) - r_j)dF_j(x) + O(g^2).$$

La idea de la prueba es que si $I_j(\cdot)$ no es de la forma CTEDSV, entonces se puede construir una función $g(\cdot)$ tal que $L_j(g) > 0$. Si $I_j(\cdot)$ no da CTEDSV, entonces existen x_1 y x_2 tales que $I_j(x_1)$ e $I_j(x_2)$ son positivos e $I_j(x_1) - x_1 < I_j(x_2) - x_2$ (o $I_j(x_1) - x_1 > I_j(x_2) - x_2$ que es similar). Construimos $g(\cdot)$ apenas positiva alrededor de x_1, apenas negativa alrededor de x_2 y cero fuera de $[x_1, x_2]$, asegurándonos que $\int_0^\infty g(x)dF_j(x) = 0$.

Debido a que u' es decreciente,

$$u'(C - x_1 + I_j(x_1) - r_j) > u'(C - x_2 + I_j(x_2) - r_j).$$

Consecuentemente,

$$L_j(g) = \int_{x_1}^{x_1+\delta} g(x)u'(C - x + I_j(x) - r_j)dF_j(x)$$

$$+\int_{x_2}^{x_2+\delta} g(x)u'(C-x+I_j(x)-r_j)dF_j(x)+O(g^2)>0,$$

contradiciendo la optimalidad de $I_j(\cdot)$.

Esta prueba debe modificarse en el modelo multiperíodo debido a las dos razones siguientes:

(1) $I_j(\cdot)$ e $\bar{I}_j(\cdot)$ no necesariamente tienen el mismo valor actuarial puesto que los valores críticos correspondientes a estos contratos no necesariamente son iguales. Por lo que la r_j, que aparece en el primer término de [?], puede ser distinta en cada argumento de u.

(2) En el modelo multiperíodo se tienen que maximizar las $W_i, i=1,...,J$, y las W_i son las soluciones del sistema de ecuaciones [?]. Pero, cualquier perturbación de $I_j(\cdot)$ cambiará no sólo W_j sino también W_i, $i \neq j$. Por lo que mostrar que $L_j(g)>0$ no es suficiente para contradecir la optimalidad de $I_j(\cdot)$.

Las compañías de seguros usualmente ofrecen contratos que especifican el deducible, pero no los valores críticos. La razón es que los valores críticos están endógenamente determinados, por lo que se prefiere una definición del contrato que no haga alusión a ellos. A los contratos de cobertura total por encima de un deducible, que no especifican la relación de x con el valor crítico, los llamaremos CTED. El siguiente Teorema es el resultado principal de esta sección. Aquí mostramos que $(I_j(x)-x)$ sería lo mismo para todos los daños, x, que excedan el valor crítico.

Teorema 4.4.1 (3.c.1) *Si en cualquier categoría j, el asegurado puede comprar cualquier contrato $I_j(\cdot)$ al precio de [?], entonces el asegurado prefiere el contrato que le proporcione CTEDSV, y siempre existe un contrato óptimo que otorgue CTED.*

Demostración. Supongamos que cuando nos encontramos en la categoría i, el contrato óptimo es $I_i(x)$, el valor crítico óptimo es y, y la utilidad esperada correspondiente es $W_i, i=1,...,J$. Supongamos que para alguna categoría j, el contrato óptimo $I_j(\cdot)$ no es de la forma CTEDSV. En este caso existen $x_1, x_2 > y$ tales que $I_j(x_1)>0, I_j(x_2)>0$ y $(I_j(x_1)-x_1)<(I_j(x_2)-x_2)$. Definimos el contrato $\bar{I}_j(x)=I_j(x)+g(x)$, donde $g(x)$ está definida como sigue:

$$g(x)=\begin{cases} p_2\varepsilon & si\ x_1 \leq x \leq x_1+\delta \\ -p_1\varepsilon & si\ x_2 \leq x \leq x_2+\delta \\ 0 & \text{en otro caso} \end{cases}, \qquad (4.4.11)$$

donde p_k denota la probabilidad de que el daño se encuentre en el intervalo $(x_k, x_k+\delta), k=1,2$, y ε y δ son números suficientemente pequeños. La función $g(x)$ fue construida de tal manera que para un valor crítico fijo y_j, los valores actuariales de $I_j(x)$ e $\bar{I}_j(x)$ sean los mismos.

Supongamos ahora que si nos encontramos en la categoría $i \neq j$, el comprador adquiere $I_i(x)$ y cuando esté en j adquiera $\bar{I}_j(x)$, y denotamos las utilidades esperadas máximas que se pueden obtener mediante esta estrategia como $\bar{W}_i$. En lo que sigue se demuestra que $\bar{W}_i \geq W_i$ para $i = 1, ..., J$ y $\bar{W}_k > W_k$ para alguna k, contradiciendo la suposición de que las $I_i(x)$ son óptimas e implicando que ningún contrato puede ser óptimo a menos que otorgue CTEDSV.

Sean $\hat{W}_i, i = 1, ..., J$, las utilidades esperadas del comprador si obtiene $I_i(x)$ cuando se encuentre en $i \neq j$ e $\bar{I}_j(x)$ cuando esté en j, y si mantiene los valores críticos no óptimos y_j que se determinan al comprar $I_i(x)$. Se sigue entonces de la no optimalidad de y_j que $\bar{W}_i \geq \hat{W}_i, i = 1, ..., J$. Puesto que mostramos en el [?] que $\hat{W}_i \geq W_i$ para $i = 1, ..., J$, y $\hat{W}_k > W_k$ para alguna k, se sigue que $\bar{W}_i \geq W_i$ para $i = 1, ..., J$ y $\hat{W}_k > W_k$ para alguna k, lo que contradice la optimalidad de $I_j(\cdot)$.

Para mostrar que siempre existe un contrato óptimo de la forma CTED, notamos que la forma del contrato óptimo $I_j(x)$ es arbitraria para $x \leq y_j$. Por lo que, si $I_j(x) = x - m_j$ para $x \geq y_j$, escogeríamos un contrato $I_j^*(x)$ CTED óptimo como sigue:

$$I_j^*(x) = \begin{cases} x - m_j \text{ para } x \geq m_j \\ 0 \text{ para } x < m_j \end{cases}.$$

$I_j^*(x)$ es equivalente a $I_j(x)$ tanto para el comprador como para la aseguradora, por lo que es una indemnización óptima también. ■

Lema 4.4.2 (3.c.1) $\hat{W}_i \geq W_i$, $i = 1, ..., J$, *y* $\hat{W}_k > W_k$ *para alguna* k.

Demostración. En [?] vimos que W_j satisface

$$W_j = \int_0^{y_j} [u(\phi_j(x)) + \beta W_{j-1}] dF_j(y) + \int_{y_j}^{\infty} [u(\psi_j(x)) + \beta W_{j+1}] dF_j(y),$$

donde $\phi_j(x) = C - r_j(I_j) - x$ y $\psi_j(x) = C - r_j(I_j) - x + I_j(x)$.

De la misma manera, si el comprador adquiere el contrato $\bar{I}_j(x)$ y mantiene el valor crítico no óptimo y_j, entonces $\hat{W}_j$satisface

$$\hat{W}_j = \int_0^{y_j} [u(\phi_j(x)) + \beta \hat{W}_{j-1}] dF_j(y) + \int_{y_j}^{\infty} [u(\bar{\psi}_j(x)) + \beta \hat{W}_{j+1}] dF_j(y), \qquad (4.4.12)$$

donde $\phi_j(x) = C - r_j(\bar{I}_j) - x$, y $\bar{\psi}_j(x) = C - r_j(\bar{I}_j) - x + \bar{I}_j(x)$. De la definición de $g(\cdot)$ en [?], se observa que $\bar{\psi}_j(x)$ y $\psi_j(x)$ son iguales fuera de los intervalos $[x_k, x_k + \delta], k = 1, 2$; y $\phi_j(x) = \bar{\phi}_j(x)$ para toda x. Por lo que sumando y restando

$$\int_{x_k}^{x_k + \delta} u\left[\psi_j(x)\right] dF_j(x), k = 1, 2;$$

al segundo término de el lado derecho de [?]

$$\hat{W}_j = \int_0^{y_j} [u(\phi_j(x)) + \beta \hat{W}_{j-1}] dF_j(y) + \int_{y_j}^{\infty} [u(\bar{\psi}_j(x)) + \beta \hat{W}_{j+1}] dF_j(y)$$

$$+\sum_{k=1}^{2} \int_{x_k}^{x_k+\delta} u\left[\psi_j(x)\right] dF_j(x) - \sum_{k=1}^{2} \int_{x_k}^{x_k+\delta} u\left[\psi_j(x)\right] dF_j(x),$$

obtenemos

$$\hat{W}_j = \int_0^{y_j} [u(\phi_j(x)) + \beta \hat{W}_{j-1}] dF_j(y) + \int_{y_j}^{\infty} [u(\psi_j(x)) + \beta \hat{W}_{j+1}] dF_j(y) + K_j, \quad (4.4.13)$$

donde

$$K_j = \sum_{k=1}^{2} \int_{x_k}^{x_k+\delta} u\left[\bar{\psi}_j(x) - \psi_j(x)\right] dF_j(x). \quad (4.4.14)$$

De la concavidad de u, de que $L_j(g) = \int_0^{\infty} g(x)u'(C - x + I_j(x) - r_j)dF_j(x) + O(g^2)$ y puesto que $I_j(x_1) - x_1 < I_j(x_2) - x_2$, se sigue que $K_j > 0$. Como en $i \neq j$ el asegurado compra el contrato $I_i(x)$, las $\hat{W}_i$ satisfacen

$$\hat{W}_i = \int_0^{y_i} \left[u(\phi_i(x)) + \beta \hat{W}_{i-1}\right] dF_i(y) + \int_{y_i}^{\infty} \left[u(\psi_i(x)) + \beta \hat{W}_{i+1}\right] dF_i(y) + K_i, \quad (4.4.15)$$

donde $K_i = 0$ para $i \neq j$.

Definiendo $D_i = \hat{W}_i - W_i, i = 1, ..., J$, $D_0 = D_1$, $D_{J+1} = 0$,podemos rescribir [?] y [?] como

$$\hat{W}_i = D_i + W_i = \left\{ \int_0^{y_i} [u(\phi_i(x)) + \beta W_{i-1}] dF_i(y) + \int_{y_i}^{\infty} [u(\psi_i(x)) + \beta W_{i+1}] dF_i(y) \right\}$$

$$+\beta D_{i-1} F_i(y_i) + \beta D_{i+1} G_i(y_i); \quad i = 1, ...J;$$

donde $G_i(y_i) = 1 - F_i(y_i)$ y

$$D_i = \beta D_{i-1} F_i(y_i) + \beta D_{i+1} G_i(y_i) + K_i, i = 1, ..., J. \quad (4.4.16)$$

Esto último se deduce de

$$D_i = \hat{W}_i - W_i = \int_0^{y_i} \left[u(\phi_i(x)) + \beta \hat{W}_{i-1}\right] dF_i(y) + \int_{y_i}^{\infty} \left[u(\psi_i(x)) + \beta \hat{W}_{i+1}\right] dF_i(y) + K_i$$

$$- \left\{ \int_0^{y_i} [u(\phi_i(x)) + \beta W_{i-1}] dF_i(y) + \int_{y_i}^{\infty} [u(\psi_i(x)) + \beta W_{i+1}] dF_i(y) \right\} =$$

$$\int_0^{y_i} \beta \left(\hat{W}_{i-1} - W_{i-1} \right) dF_i(x) + \int_{y_i}^{\infty} \beta \left(\hat{W}_{i+1} - W_{i+1} \right) + K_i = \beta D_{i-1} \int_0^{y_i} dF_i(x) +$$

$$\beta D_{i+1}\int_{y_i}^{\infty} dF_i(x) + K_i = \beta D_{i-1}F_i(y_i) + \beta D_{i+1}G_i(y_i) + K_i, \quad i = 1, ..., J.$$

Este sistema de J ecuaciones con J incógnitas (las D_j) se pueden escribir de forma matricial como

$$\mathbf{B}\cdot\mathbf{D} = \mathbf{K} \tag{4.4.17}$$

donde $\mathbf{D} = (D_1, ..., D_J)', \mathbf{K} = (K_1, ..., K_J)'$ y la matriz $\mathbf{B}_{JxJ}$ es definida por

$$B = \begin{bmatrix} (1-\beta F_1) & -\beta G_1 & & & & & \dots & 0 \\ -\beta F_2 & 1 & -\beta G_2 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & -\beta F_j & 1 & -\beta G_j & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & \ddots & \ddots & \ddots & \\ 0 & \cdots & & & & & -\beta F_J & 1 \end{bmatrix},$$

donde $F_i = F_i(y_i)$ y $G_i = G_i(y_i)$. De la matriz anterior notamos que todos los elementos de la diagonal de B y todas las sumas de los renglones son positivas, y que los elementos fuera de la diagonal son no negativos. Por lo tanto la matriz B^{-1} es una matriz que contiene solo elementos no negativos (ver el apéndice). Puesto que $K_j > 0$ se sigue de la ecuación en [?] y de que $\mathbf{D} = \mathbf{B}^{-1}\mathbf{K}$, que $D_k > 0$ para alguna k. ■

Habiendo establecido que siempre existe un contrato óptimo que otorgue cobertura total por encima de un deducible, procederemos a determinar el deducible óptimo y por lo tanto el contrato óptimo. Para ello demostraremos el siguiente teorema:

Teorema 4.4.3 (3.c.2) *Supongamos que el comprador adquiere un contrato óptimo que otorga cobertura total por encima de un deducible. Entonces los deducibles óptimos m_j^* satisfacen $m_j^* > 0, j = 1, ..., J.$*

Demostración. Sean m_j^*, y_j^* y W_j^* las notaciones del deducible óptimo, el valor crítico y la utilidad esperada resultante, respectivamente, $j = 1, ..., J$. Estos valores son obtenidos al resolver los siguientes problemas de maximización:

$$\begin{aligned} W_j = \max_{m_j}\max_{y_j}\int_0^{y_j} &\left(u\left(C - r_j - x\right) + \beta W_{j-1}\right) dF_j\left(x\right) \\ &+ \int_{y_j}^{\infty}\left(u\left(C - r_j - m_j\right) + \beta W_{j+1}\right) dF_j\left(x\right), \\ j &= 1, ..., J, \end{aligned} \tag{4.4.18}$$

sujeto a

$$r_j = (1+\lambda_j)\int_{y_j}^{\infty}(x-m_j)\,dF_j(x).$$

Empleando el hecho de que si alguna función $g(Z_1,...,Z_n)$ es maximizada por $(Z_1^*,...,Z_n^*)$, entonces Z_1^* maximiza $g(Z_1,Z_2^*,...,Z_n^*)$, podemos verificar que m_j^* es la solución del siguiente problema de maximización:

$$\max_{m_j} H(m) = \int_0^{y_j^*}\left(u(C-r_j-x)+\beta W_{j-1}^*\right)dF_j(x)+\int_{y_j^*}^{\infty}\left(u(C-r_j-m_j)+\beta W_{j+1}^*\right)dF_j(x),$$

sujeto a

$$r_j = (1+\lambda_j)\int_{y_j}^{\infty}(x-m_j)\,dF_j(x);$$

y dado $W_i^*, y_i^*, i=1,...,J$, y $m_i^*, i\neq j$.

Notando que

$$\begin{aligned}\frac{\partial}{\partial m}r_j &= \frac{\partial}{\partial m}(1+\lambda_j)\int_{y_j}^{\infty}(x-m)\,dF_j(x)\\ &= (1+\lambda_j)\left[\frac{\partial}{\partial m}\left(\int_{y_j}^{\infty}x dF_j(x)-m\int_{y_j}^{\infty}dF_j(x)\right)\right]\\ &= (1+\lambda_j)\left(-\int_{y_j}^{\infty}dF_j(x)\right) = -(1+\lambda_j)G_j(y_j),\end{aligned}$$

con $G_j(y_j)=1-F_j(y_j)$, se obtiene

$$\frac{\partial}{\partial m}H(m)$$

$$= \frac{\partial}{\partial m}\left[\int_0^{y_j^*}\left(u(C-r_j-x)+\beta W_{j-1}^*\right)dF_j(x)+\int_{y_j^*}^{\infty}\left(u(C-r_j-m)+\beta W_{j+1}^*\right)dF_j(x)\right]$$

$$\begin{aligned}= &\int_0^{y_j^*}u'(C-r_j-x)(1+\lambda_j)G_j\left(y_j^*\right)dF_j(x)\\ &+\int_{y_j^*}^{\infty}u'(C-r_j-m)\left(G_j\left(y_j^*\right)+\lambda_j G_j\left(y_j^*\right)-1\right)dF_j(x)\\ = &\ G_j\left(y_j^*\right)\\ &\left\{\int_0^{y_j^*}u'(C-r_j-x)\,dF_j(x)+\lambda_j\left[\begin{array}{c}\int_0^{y_j^*}u'(C-r_j-x)\,dF_j(x)\\ +\int_{y_j^*}^{\infty}u'(C-r_j-m)\,dF_j(x)\end{array}\right]\right\}\\ &-F_j\left(y_j^*\right)\int_{y_j^*}^{\infty}u'(C-r_j-m)\,dF_j(x).\end{aligned}$$

Como $u'(C - r_j - m)$ no depende de x (aunque r_j este en términos de x, pero ya ha tomado un valor), lo podemos cambiar de integral en el último término anterior, para obtener:

$$\begin{aligned}
& F_j(y_j^*) \int_{y_j^*}^{\infty} u'(C - r_j - m)\, dF_j(x) \\
= & \int_0^{y_j^*} dF_j(x) \left[\int_{y_j^*}^{\infty} u'(C - r_j - m)\, dF_j(x) \right] \\
= & \int_0^{y_j^*} dF_j(x) \left[u'(C - r_j - m) \int_{y_j^*}^{\infty} dF_j(x) \right] \\
= & \int_0^{y_j^*} u'(C - r_j - m)\, dF_j(x) \left[\int_{y_j^*}^{\infty} dF_j(x) \right] \\
= & \int_{y_j^*}^{\infty} dF_j(x) \int_0^{y_j^*} u'(C - r_j - m)\, dF_j(x) \\
= & G_j(y_j^*) \int_0^{y_j^*} u'(C - r_j - m)\, dF_j(x).
\end{aligned}$$

Por lo que agrupando nos queda

$$\frac{\partial}{\partial m} H(m) = G_j(y_j^*) \{ \int_0^{y_j^*} \left[u'(C - r_j - x) - u'(C - r_j - m) \right] dF_j(x)$$

$$+\lambda_j \left[\int_0^{y_j^*} u'(C - r_j - x)\, dF_j(x) + \int_{y_j^*}^{\infty} u'(C - r_j - m)\, dF_j(x) \right] \}.$$

Debido a que $u'(\cdot)$ es convexa decreciente positiva ($u'(C - r_j - x) - u'(C - r_j) > 0$), se sigue que $\frac{\partial}{\partial m} H(m) > 0$ en $m = 0$, y por lo tanto el deducible óptimo m_j^* debe ser estrictamente positivo. ■

Este resultado difiere a los contratos de un solo período, porque en ellos, el deducible positivo es necesario solo si $\lambda > 0$ y en el multiperíodo $\frac{\partial H}{\partial m} |_{m=0} > 0$, aun si $\lambda_j = 0$. La razón de esta diferencia se debe a que en los modelos de un solo período, realizar un reclamo no involucra ningún costo, por lo que si la prima no es muy costosa, el asegurado prefiere asegurarse lo más posible, y por lo tanto bajar el deducible a cero. En el caso multiperiodico, realizar un reclamo nos mueve a una categoría de mayor riesgo, y nos costaría más que pagar una prima mayor que nos bajara el deducible, por lo que el asegurado prefiere un deducible positivo.

4.5 Apéndice

Lema 4.5.1 (A.1) *Sea A una matriz de nxn, tales que todos los elementos de la diagonal y todas las sumas de los renglones son positivos, y todos los elementos fuera de la diagonal no positivos. Entonces la matriz A^{-1} contiene solamente elementos no negativos*

Demostración. Para encontrar A^{-1} empezamos multiplicando a la matriz A por la matriz

$$\begin{bmatrix} \frac{1}{a_{11}} & 0 & \dots & 0 \\ 0 & \frac{1}{a_{22}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{a_{nn}} \end{bmatrix},$$

de donde obtenemos la matriz similar

$$A = B = \begin{bmatrix} 1 & b_{12} & \dots & b_{1n} \\ b_{21} & 1 & & b_{2n} \\ \vdots & & \ddots & \vdots \\ b_{n1} & & & 1 \end{bmatrix}.$$

Notamos aquí que las características de la matriz siguen siendo las mismas, es decir, la suma de los renglones es positiva $\sum_{j=1}^{n} b_{ij} = \sum_{j=1}^{n} \frac{a_{ij}}{a_{ii}} = \frac{1}{a_{ii}} \sum_{j=1}^{n} a_{ij} > 0, i = 1, .., n;$ los elementos de la diagonal también son uno, que es positivo, y los elementos fuera de la diagonal no positivos $b_{ij} = \frac{1}{a_{ii}}(a_{ij}) \leq 0$, para $i \neq j$. Además notamos que como $\sum_{j=1}^{n} b_{ij} = 1 + \sum_{j \neq i}^{n} b_{ij} > 0$ y de que la suma de los elementos de la diagonal es no positiva, $0 \geq \sum_{j \neq i}^{n} b_{ij} > -1$.

Para empezar a obtener ceros debajo de la diagonal, empezamos por la primer columna, por lo que multiplicamos al primer renglón por $-b_{i1}$ y se lo sumamos al renglón i, es decir, en forma vectorial tendríamos,

$$-b_{i1} \begin{pmatrix} 1 & b_{12} & .. & b_{1n} \end{pmatrix} + \begin{pmatrix} b_{i1} & b_{i2} & \dots & 1 & \dots & b_{in} \end{pmatrix} =$$

$$\begin{pmatrix} 0 & b_{i2} - b_{i1}b_{12} & \dots & 1 - b_{i1}b_{1i} & \dots & b_{in} - b_{i1}b_{1n} \end{pmatrix}.$$

Sea $b_{ij} - b_{i1}b_{1j}$ un elemento de este renglón que no es de la diagonal, es decir, $i \neq j$. Para ver en que intervalo se encuentra este número, notamos que $-1 < b_{ij} \leq 0$, $i \neq j$, así que $-1 < b_{1j} \leq -b_{i1}b_{1j} \leq 0$.

Ahora bien, si sumo b_{ij} a los elementos de la desigualdad y notando que $-1 < b_{ij} + b_{1j}$ porque ambos elementos pertenecen al mismo renglón, tenemos

$$-1 < b_{ij} + b_{1j} \leq b_{ij} - b_{i1}b_{1j} \leq 0.$$

Para ver el signo del elemento de la diagonal, como $-1 < -b_{i1}b_{1i} \leq 0$, y sumando1 a cada término, nos queda

$$0 > 1 - b_{i1}b_{1i} \geq 1.$$

Hasta aquí el elemento de la diagonal es mayor que cero y los que están fuera de la diagonal son no positivos, falta ver como queda la suma del renglón. Como suma tenemos, $\sum_{j=1}^{n} b_{ij} - b_{i1}b_{1j} = \sum_{j=1}^{n} b_{ij} - b_{i1}\sum_{j=1}^{n} b_{1j}$, y como $\sum_{j=1}^{n} b_{ij} \geq 0, \forall i = 1, .., n$ y $-b_{i1} \geq 0$, entonces $\sum_{j=1}^{n} b_{ij} - b_{i1}\sum_{j=1}^{n} b_{1j} > 0$.

Así, el renglón vuelve a ser igual que como la hipótesis inicial, por lo que si multiplicamos a la matriz B por la matriz

$$\begin{bmatrix} 1 & 0 & \dots & & \dots & 0 \\ 0 & \frac{1}{1-b_{21}b_{12}} & & & & \vdots \\ \vdots & \dots & \ddots & & & \\ 0 & 0 & 0 & \frac{1}{1-b_{i1}b_{1i}} & \dots & 0 \\ \vdots & & & & \ddots & \\ 0 & \dots & & & & \frac{1}{1-b_{n1}b_{1n}} \end{bmatrix},$$

obtenemos una matriz como cuando multiplicamos la matriz A por la matriz con término en la diagonal igual a $\frac{1}{a_{ii}}$,salvo que debajo del primer uno, tenemos el mismo valor cero en todas las entradas

$$\begin{bmatrix} 1 & b_{12} & \dots & & b_{1n} \\ 0 & 1 & b'_{23} & \dots & b'_{2n} \\ 0 & b'_{32} & 1 & \dots & b'_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & b'_{n2} & \dots & & 1 \end{bmatrix},$$

donde $b'ij = \frac{b_{ij}-b_{i1}b_{1j}}{1-b_{i1}b_{1i}}$, $\forall i = 2, ..., n$ y $i \neq j$. Notamos que la submatriz obtenida al quitar el primer renglón y la primer columna de B, cumple la hipótesis de B nueva-

mente, por lo que si repetimos el proceso para las siguientes submatrices, obtenemos

$$B = \begin{bmatrix} 1 & b_{12} & \dots & & & b_{1n} \\ 0 & 1 & b'_{23} & \dots & & b'_{2n} \\ 0 & \dots & 1 & b^{(2)}_{34} & \dots & b^{(2)}_{3n} \\ & \ddots & & \ddots & & \vdots \\ 0 & \dots & & 0 & 1 & b^{(n-2)}_{n-1,n} \\ 0 & \dots & & & & 1 \end{bmatrix}.$$

Con el propósito de mejorar la notación hacemos $c_{ij} = b^{(k)}_{n-1,n}, \forall k = 1, ..., n-2$ y $(1) = \prime$, e $i < j, i = 1, ..., n$. Para hacer los elementos c_{ij} iguales a cero, partimos de la última columna. Para ello, multiplicamos el último renglón por $-c_{n-1,n}$ y se lo sumamos al renglón anterior, y así de manera escalonada hacia arriba, logramos hacer a la matriz B como la identidad. Por último notamos que la matriz A fue multiplicada solamente por matrices elementales positivas, es decir por matrices del tipo

1) multiplicar y sumar un número no negativo de un renglón a otro

$$\begin{bmatrix} 1 & 0 & \dots & & & & & 0 \\ 0 & 1 & 0 & & & & & 0 \\ \vdots & & \ddots & \ddots & & & & \vdots \\ 0 & \dots & 0 & -b_{ij} & 1 & & & \\ \vdots & & & \ddots & 1 & 0 & \dots & 0 \\ & & & & & \ddots & & \\ 0 & \dots & & & & & & \\ & & & & & & & 1 \end{bmatrix} \text{ ó } \begin{bmatrix} 1 & 0 & \dots & & & & & 0 \\ 0 & 1 & 0 & & & & & 0 \\ \vdots & & \ddots & \ddots & & & & \vdots \\ 0 & \dots & 1 & -c_{ij} & 0 & & & \\ \vdots & & & \ddots & 1 & 0 & \dots & 0 \\ & & & & & \ddots & & \\ 0 & \dots & & & & & & \\ & & & & & & & 1 \end{bmatrix},$$

y

2) de hacer los elementos de la diagonal igual a 1

$$\begin{bmatrix} \frac{1}{a_{11}} & 0 & \dots & 0 \\ 0 & \frac{1}{a_{22}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{a_{nn}} \end{bmatrix},$$

donde $-b_{ij}, -c_{ij}$, con $i \neq j$ son no negativos y $\frac{1}{a_{ii}} > 0$. ■

Capítulo 5

Determinación de la prima óptima cuando se desconoce la distribución de los reclamos

5.1 Introducción

En el primer capítulo de este trabajo, hicimos una presentación inicial de este problema, explicando cuál era el objetivo y cuáles son los elementos del modelo de decisión secuencial correspondiente. En este capítulo nos proponemos obtener una estrategia óptima para determinar la prima de un seguro cuando se desconocen los parámetros de la distribución de los reclamos y de su monto. Como vimos desde el primer capítulo, recurriremos a principios básicos de estadística Bayesiana para desarrollar un proceso de aprendizaje sobre las distribuciones desconocidas con base en el cual será determinada la prima al inicio de cada período, y por ello, en la sección 2 expondremos brevemente los elementos de estadística Bayesiana que emplearemos más adelante .

5.2 Elementos de estadística y de decisión Bayesiana

Empezaremos por recordar de manera muy general que tipos de problemas resuelve la estadística, para después ver su relación con la estadística bayesiana.Un problema de estadística en general, es un problema en el cual se han de analizar los datos extraídos aleatoriamente de una población (a los que llamaremos muestra aleatoria de la población), cuyos elementos toman valores de acuerdo a una distribución de probabilidad desconocida y en el que se debe realizar algún tipo de inferencia acerca de tal distribución. En otras palabras, en un problema de estadística existen dos o más distribuciones de probabilidad que podrían haber generado algunos datos experimen-

tales y a partir de una muestra aleatoria, se necesita inferir la distribución que pudo haberlos generado. En la mayoría de los problemas estadísticos, lo que hace que se desconozca tal distribución de probabilidad, es que se desconocen los valores de uno o más parámetros de esta distribución, y en términos generales, el problema de la inferencia estadística consiste en determinar dónde es probable que se encuentre el verdadero valor del parámetro θ en el espacio paramétrico Θ, partiendo de ciertas observaciones de la función $f(x \mid \theta)$. A θ lo denominaremos el *estado de naturaleza.*

En estadística Bayesiana, a diferencia de la estadística clásica, antes de disponer de las observaciones de $f(x \mid \theta)$, el experimentador tiene cierta información, conocimientos o creencias acerca de dónde es probable que se encuentre el verdadero valor de θ, y para ello construye una distribución de probabilidad de θ en el conjunto Θ. Esta distribución se denomina distribución *a priori* de θ porque se establece antes de obtener observaciones de $f(x \mid \theta)$ y la denotaremos como $p(\theta)$.

A medida que se va obteniendo información de las observaciones, la distribución a priori se modifica usando el Teorema de Bayes, para obtener una nueva distribución a la que llamaremos *a posteriori.*

Un problema de decisión estadística (en donde se utiliza comúnmente la estadística Bayesiana) es aquel donde después de haber analizado los datos experimentales, se debe tomar una decisión o acción, dentro de una clase disponible de acciones, con la propiedad de que las consecuencias de cada acción disponible dependen de la estimación de cierto parámetro. En la teoría de decisión Bayesiana, la consecuencia corresponde al valor de la recompensa que esperamos obtener, y ésta será mayor en la medida en que nuestra decisión se halla tomado con la mejor estimación del parámetro.

La elección misma de una inferencia (más allá del resumen de datos) puede verse como un problema de decisión, donde el espacio de acciones es el conjunto de todos los posibles enunciados de inferencia y se cuenta con una función de recompensas que reflejan el éxito o el fracaso del conocimiento usado. Si se elige una acción particular a_1 y θ_1 resulta ser el verdadero estado de naturaleza, entonces se obtendría una recompensa $r(\theta_1, a_1)$.

5.2.1 Distribución a priori

Como mencionamos antes, la manera usual de referirnos a la información a priori es en términos de la distribución de probabilidad en Θ que se construye a través de diversos métodos, y refleja la información y conocimientos previos del experimentador acerca de donde es probable que se encuentre el verdadero valor de θ en el espacio paramétrico Θ, antes de disponer de observaciones de $f(x \mid \theta)$. La información a priori sobre θ, a menudo es muy precisa, por lo que es natural establecer creencias a priori en términos

de probabilidades de los posibles valores verdaderos de θ. Así, la densidad tiene como "variable aleatoria " al parámetro θ y se establece como

$$P(\theta \in A) = \int_A dF(\theta) = \begin{cases} \int_A p(\theta)\, d\theta & \text{en el caso absolutamente continuo} \\ \sum_{\theta \in A} p(\theta) & \text{en el caso discreto} \end{cases} ;$$

Aunque se hable de probabilidades respecto a θ, en realidad no hay nada de "aleatorio" respecto a θ, puesto que la designación de este valor se basa en las "probabilidades personales" que reflejan el grado de creencias personales del experimentador en la verosímilitud del parámetro.

Ejemplo 5.2.1 *Sea θ la probabilidad de obtener un águila cuando se lanza cierta moneda y supóngase que se sabe que la moneda es equilibrada o tiene un águila de un sólo lado. Por tanto, los únicos valores posibles de θ son $\theta = \frac{1}{2}$ si la moneda es equilibrada y $\theta = 1$ si no lo es; y claramente la distribución de θ es Bernoulli. Si la probabilidad a priori de que la moneda sea equilibrada es q, entonces la función de probabilidad a priori de θ es $p\left(\theta = \frac{1}{2}\right) = q$ y $p(\theta = 1) = 1 - q$.*

5.2.2 Métodos para la determinación subjetiva de la densidad a priori

Si Θ es discreto se determina la probabilidad subjetiva de cada elemento (como en el ejemplo anterior), pero si es continuo el problema se complica considerablemente.

a. La aproximación por histograma Cuando Θ es un intervalo de la línea real, la aproximación más obvia es con el histograma. Se divide Θ en intervalos, se determina la probabilidad subjetiva correspondiente a cada intervalo, y se gráfica un histograma de probabilidad. El qué tan detallado se tenga que hacer el histograma, depende de consideraciones de robustez. Con base en el histograma se puede trazar la densidad $p(\theta)$. El método es muy sencillo, aunque se tienen varios problemas al trabajar con la densidad obtenida en esta forma como el que en cierto intervalo o varios de ellos, no se hayan considerado desviaciones importantes de la densidad, o que la densidad en si no tenga "colas".

b.- La aproximación por verosimilitud relativa

Esta aproximación también se usa cuando Θ es un subconjunto de la línea real. Consiste simplemente de comparar las "verosimilitudes" relativas de varios puntos en Θ, y trazar directamente la densidad a priori de estas determinaciones.

Ejemplo 5.2.2 *Sea $\Theta = [0, 1]$. Una buena idea es empezar a determinar las verosimilitudes relativas de los puntos parametrales "más probables" y "menos probables". Supongamos que el punto parametral $\theta = \frac{3}{4}$ se siente que es el más probable, mientras que*

$\theta = 0$ es el menos probable. También, $\theta = \frac{3}{4}$ se estima tres veces más probable que el valor de 0. Es suficiente determinar las verosimilitudes relativas de tres puntos más, por ejemplo, $\frac{1}{4}, \frac{1}{2}$ y 1. Por simplicidad, todos los puntos se comparan con $\theta = 0$. Se decide que $\theta = \frac{1}{2}$ y $\theta = 1$ son doblemente probables que $\theta = 0$. Se le asigna al punto base $\theta = 0$ el valor 1 de la densidad a priori..

El principal problema se encuentra cuando Θ es no acotado, pues las verosimilitudes relativas se determinan solo en regiones finitas, aunque este es el método mayormente empleado.

c.-Coincidiendo con una forma funcional dada. Esta aproximación es la más usada para determinar la densidad a priori. La idea es simplemente asumir que $p(\theta)$ es de una forma funcional dada, y escoger los parámetros de la densidad que más coincida con las creencias a priori. Por ejemplo, si la a priori se asume que tiene una forma funcional $\mathfrak{Be}(\alpha, \beta)$, se puede estimar la media a priori, μ, y varianza, σ^2, y usar las relaciones $\mu = \frac{\alpha}{(\alpha+\beta)}$, $\sigma^2 = \frac{\alpha\beta}{[(\alpha+\beta)^2(\alpha+\beta+1)]}$ para determinar α y β (método de momentos).

La dificultad de este método estriba en que las "colas" de la densidad pueden tener un efecto drástico en sus momentos o que la densidad simplemente carezca de ellos.

5.2.3 Distribución a posteriori

El análisis Bayesiano se desarrolla para combinar la información a priori y la información muestral en la determinación de la llamada distribución a posteriori de θ dado $\mathbf{x}$. Con base en esta última, se realizan inferencias o se toman decisiones.

Respecto a la distribución a posteriori, supongamos que las variables aleatorias $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ constituyen una muestra aleatoria de una distribución cuya función de probabilidad es $f(x \mid \theta)$. Supongamos también que el parámetro θ es desconocido y que la función de probabilidad a priori de θ es $p(\theta)$.

Puesto que las variables aleatorias $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ son idénticamente distribuidas, con función de densidad $f(x \mid \theta)$, la distribución de probabilidad conjunta $f(x_1, ..., x_n \mid \theta)$ está dada por la ecuación

$$f(x_1, ..., x_n \mid \theta) = f(x_1 \mid \theta) \cdots f(x_n \mid \theta).$$

Si se utiliza la notación vectorial $\mathbf{x} = (x_1, x_2, ..., x_n)$, entonces la función de densidad conjunta se puede escribir como $f_n(\mathbf{x} \mid \theta)$.

Como se supone que el parámetro θ tiene una distribución cuya función de densidad es $p(\theta)$, la función de densidad conjunta $f_n(\mathbf{x} \mid \theta)$ es la función de densidad conjunta condicional de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ para un valor dado de θ. Si se multiplica esta densidad

conjunta condicional por $p(\theta)$, se obtiene la función de densidad conjunta $(n+1)$-dimensional de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ y θ, de la forma $f_n(\mathbf{x} \mid \theta) p(\theta)$. La función de densidad conjunta marginal de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ se puede obtener ahora integrando esta función sobre todos los valores de θ. Por tanto, la función de densidad conjunta marginal que corresponde a un vector aleatorio $n-$dimensional $m_n(\mathbf{x})$ de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ se puede escribir de la forma

$$m_n(\mathbf{x}) = \int_{\Theta} f_n(\mathbf{x} \mid \theta) p(\theta) \, d\theta.$$

Además, la función de densidad condicional de θ dado $\mathbf{X}_1 = x_1, \mathbf{X}_2 = x_2, ..., \mathbf{X}_n = x_n$, que se denota por $p(\theta \mid \mathbf{x})$, debe ser igual a la función de densidad conjunta de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ y θ dividida por la función de densidad conjunta marginal de $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$. Por ello resulta que

$$p(\theta \mid \mathbf{x}) = \frac{f_n(\mathbf{x} \mid \theta) p(\theta)}{m_n(\mathbf{x})} \text{ para } \theta \in \Theta. \tag{5.2.1}$$

La densidad sobre Θ representada por la función de densidad condicional $p(\theta \mid \mathbf{x})$ es la distribución a posteriori de θ, y, como la notación lo indica, se define como la distribución condicional de θ dada la observación muestral $\mathbf{x}$

La fórmula para $p(\theta \mid \mathbf{x})$ es una aplicación directa del Teorema de Bayes. El nombre de distribución a posteriori es indicativo del papel de $p(\theta \mid \mathbf{x})$. Tal y como $p(\theta)$ refleja las creencias de la distribución a priori de la experimentación, $p(\theta \mid \mathbf{x})$ refleja las creencias actualizadas sobre θ, después o a posteriori de la experimentación.

a. Función de verosimilitud

El denominador de la parte derecha de la ecuación 5.2.1 es simplemente la integral del numerador sobre todos los valores posibles de θ. Aunque el valor de esta integral depende de los valores observados $x_1, ..., x_n$, no depende de θ y se puede tratar como una constante cuando la parte derecha de la ecuación 5.2.1 se considera como una función de densidad de θ. Se puede remplazar entonces la ecuación 5.2.1 por la siguiente relación

$$p(\theta \mid \mathbf{x}) \propto f_n(\mathbf{x} \mid \theta) p(\theta) = h(x, \theta), \tag{5.2.2}$$

donde $\propto$ significa que las funciones son directamente proporcionales. El factor constante apropiado que establecerá la igualdad de las dos partes de la relación 5.2.2 se puede determinar en cualquier momento utilizando el hecho de que $\int_{\Theta} p(\theta \mid \mathbf{x}) \, d\theta = 1$, puesto que $p(\theta \mid \mathbf{x})$ es una densidad.

Cuando la función de densidad conjunta $f_n(\mathbf{x} \mid \theta)$ se considera como una función de θ para valores dados de $x_1, ..., x_n$, se llama función de verosimilitud. Así que 5.2.2

afirma que la función de densidad a posteriori de θ es proporcional al producto de la función de verosimilitud y la función de densidad a priori de θ.

Utilizando la relación proporcional 5.2.2, generalmente es posible determinar la función de densidad a posteriori de θ sin resolver explícitamente la integral $\int_{\Theta} f_n(\mathbf{x} \mid \theta) p(\theta) d\theta$. Si se puede reconocer la parte derecha de la relación 5.2.2 como una de las densidades más conocidas, excepto, posiblemente por una constante, entonces se puede determinar con facilidad el factor apropiado que convierte la parte derecha de 5.2.2 en una función de densidad propia de θ.

b. Familias conjugadas

En general $m(\mathbf{x})$ y $p(\theta \mid \mathbf{x})$ no son fácilmente calculables. Si, por ejemplo, X se distribuye $N(\theta, \sigma^2)$ y θ a su vez se distribuye *Cauchy* con parámetros μ y β, entonces $p(\theta \mid \mathbf{x})$ solo puede evaluarse numéricamente. Una gran parte de la literatura Bayesiana esta enfocada a encontrar distribuciones a priori para las cuales se pueda calcular fácilmente $p(\theta \mid \mathbf{x})$.

Definición 5.2.3 *Sea $\mathfrak{F}$ la clase de funciones de densidad $f(\mathbf{x} \mid \theta)$. A una clase $\mathfrak{P}$ de distribuciones a priori se le llama familia conjugada para $\mathfrak{F}$ si $p(\theta \mid \mathbf{x})$ esta en la clase $\mathfrak{P}$ para toda $f \in \mathfrak{F}$ y $p \in \mathfrak{P}$.*

Para una clase dada de densidades $\mathfrak{F}$, con frecuencia se puede determinar una familia conjugada examinando las funciones de verosimilitud $l_{\mathbf{x}}(\theta) = f(\mathbf{x} \mid \theta)$, y escogiendo, como familia conjugada, la clase de distribuciones con la misma forma funcional de esas funciones de verosimilitud. Las a priori resultantes son llamadas a priori conjugadas naturales.

Cuando se trabaja con aprioris conjugadas, no hay necesidad de calcular explícitamente $m(\mathbf{x})$. La razón es que, como $p(\theta \mid \mathbf{x}) = \frac{h(\mathbf{x},\theta)}{m(\mathbf{x})}$, los factores que involucran a θ en $p(\theta \mid \mathbf{x})$ deben ser los mismos que para los factores que involucran θ en $h(\mathbf{x}, \theta)$. Sólo es necesario observar a los factores que involucran a θ en $h(\mathbf{x}, \theta)$, y ver si estos pueden ser reconocidos como pertenecientes a una distribución en particular. Si asi es, $p(\theta \mid \mathbf{x})$ es esa distribución. La densidad marginal $m(\mathbf{x})$ puede ser determinada, si se desea, dividiendo $h(\mathbf{x}, \theta)$ por $p(\theta \mid \mathbf{x})$.

Ejemplo 5.2.4 *Supongamos que $\mathbf{X} = (X_1, X_2, ..., X_n)$ es una muestra aleatoria de una distribución Poisson. Asi que $X_i \sim \mathcal{P}(\theta), i = 1, 2, .., n$, y*

$$f(\mathbf{x} \mid \theta) = \prod_1^n \left[\frac{\theta^{x_i} e^{-\theta}}{x_i!}\right] = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_1^n [x_i!]}.$$

Aquí, $\mathfrak{F}$ es la clase de todas las densidades anteriores. Observando que la función de verosimilitud para tales densidades se asemeja a una densidad gama, una suposición plausible para una familia conjugada de las distribuciones a priori es la clase de las distribuciones gama. Por lo que asumimos $\theta \sim \mathcal{G}(\alpha, \beta)$, y observamos que

$$h(\mathbf{x}, \theta) = f(\mathbf{x} \mid \theta) p(\theta) = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_1^n [x_i!]} * \frac{\theta^{\alpha-1} e^{-\theta/\beta} I_{(0,\infty)}(\theta)}{\Gamma(\alpha)\beta^{\alpha}} = \frac{e^{-\theta(n+1/\beta)} \theta^{\left(n\bar{x}+\alpha-1\right)} I_{(0,\infty)}(\theta)}{\Gamma(\alpha)\beta^{\alpha} \prod_1^n [x_i!]}.$$

Los factores que involucran a θ en esta última expresión son claramente reconocibles como pertenecientes a la distribución $\mathcal{G}\left(n\bar{x}+\alpha, [n+1/\beta]^{-1}\right)$. Esta debe ser $p(\theta \mid \mathbf{x})$. Como la función a posteriori es una distribución gama, se sigue que la clase de las distribuciones gama son una familia conjugada (natural) para $\mathfrak{F}$.

En este ejemplo, $m(\mathbf{x})$ puede determinarse al dividir $h(\mathbf{x}, \theta)$ por $p(\theta \mid \mathbf{x})$ y cancelar los factores que involucran a θ. El resultado es

$$m(\mathbf{x}) = \frac{h(\mathbf{x}, \theta)}{p(\theta \mid \mathbf{x})} = \frac{\left(\Gamma(\alpha)\beta^{\alpha} \prod_1^n [x_i!]\right)^{-1}}{\left\{\Gamma\left(n\bar{x}+\alpha\right)[n+1/\beta]^{-\left(\alpha+n\bar{x}\right)}\right\}^{-1}}.$$

5.2.4 Regla de decisión Bayes

Para un problema de decisión múltiple se definirán $\theta_1, ..., \theta_k$ como los k valores posibles de θ, y $a_1, ..., a_m$ serán las m acciones posibles que se pueden elegir. Además, para $i = 1, ..., k$ y $j = 1, ..., m$, sea r_{ij} la ganancia obtenida por el experimentador cuando $\theta = \theta_i$ y se elige la acción a_j. Finalmente, para $i = 1, ..., k$ sea p_i la probabilidad inicial de que $\theta = \theta_i$. Por tanto, $p_i \geq 0$ y $p_1 + ... + p_k = 1$.

Si el experimentador debe elegir una de las acciones $a_1, ..., a_m$ sin poder observar ningún dato muestral relevante, entonces la recompensa esperada $r(a_j)$ de seleccionar la acción a_j será

$$r(a_j) = \sum_{i=1}^{k} p_i r_{ij}.$$

La elección de una acción para la cual la recompensa es máxima se denomina decisión Bayes.

Ejemplo 5.2.5 : *Obtención de una decisión Bayes. Considérese un problema de decisión múltiple en el cual $k = 3$ y $m = 4$, y las ganancias r_{ij} están dadas por la siguiente tabla:*

	a_1	a_2	a_3	a_4
θ_1	1	2	3	4
θ_2	3	0	1	2
θ_3	4	2	1	0

De esta tabla se deduce que las recompensas obtenidas de las cuatro acciones posibles son las siguientes:

$$r(a_1) = p_1 + 3p_2 + 4p_3$$
$$r(a_2) = 2p_1 + 2p_3$$
$$r(a_3) = 3p_1 + p_2 + p_3$$
$$r(a_4) = 4p_1 + 2p_2$$

Para cualesquiera probabilidades iniciales p_1, p_2 y p_3, una decisión Bayes se encuentra determinando simplemente la acción para la cual la recompensa es máxima. A modo de ilustración, si $p_1 = 0.5, p_2 = 0.2$ y $p_3 = 0.3$, entonces $r(a_1) = 2.3, r(a_2) = 1.6, r(a_3) = 2.0$ y $r(a_4) = 2.4$. Por tanto, a_4 es la única decisión Bayes. Si $\theta = \theta_1$, se puede observar en la primer fila de la tabla de recompensas que a_4 tiene la mayor recompensa entre las cuatro acciones. Por tanto, si la probabilidad inicial p_1 esta suficientemente cerca de 1, entonces a_4 será la decisión Bayes. Análogamente, si $\theta = \theta_2$, entonces a_1 tendrá la mayor recompensa entre las cuatro acciones. Por tanto si p_2 esta suficientemente cerca de 1, entonces a_1 será la decisión Bayes. Finalmente, si $\theta = \theta_3$, entonces a_1 tiene la mayor recompensa entre las cuatro acciones. Por tanto, si p_3 esta suficientemente cerca de 1, entonces a_1 será la decisión Bayes. Se determinará ahora si existen probabilidades iniciales p_1, p_2 y p_3, para las cuales a_3 sea una decisión Bayes.

Los siguientes resultados se pueden obtener de las ecuaciones anteriores: $r(a_2) < r(a_3)$ si y solo si, $p_1 + p_2 > p_3$, y $r(a_4) < r(a_3)$ si y solo si, $p_1 + p_2 < p_3$. Por tanto, la única condición para que a_3 pudiera ser una decisión Bayes es que $p_1 + p_2 = p_3$. Pero si $p_1 + p_2 = p_3$, entonces se deduce que $p_1 + p_2 = \frac{1}{2}$ y $p_3 = \frac{1}{2}$, y se puede verificar de las ecuaciones que $r(a_2) = r(a_3) = r(a_4) = 1 + 2p_1$ y $r(a_1) = \frac{5}{2} + 2p_2 > 1 + 2p_1$. De esto último se puede concluir que como $r(a_1)$ es mayor estrictamente que $r(a_3)$, $r(a_3)$ no puede maximizarse y a_3 no puede ser una decisión Bayes.

Considerese ahora un problema de decisión múltiple general sujeto a las siguientes condiciones: Existen k valores posibles del parámetro θ, hay m acciones posibles, la recompensa que resulta de elegir la acción a_j cuando $\theta = \theta_i$ es r_{ij} para $i = 1, ..., k$ y $j = 1, ..., m$, y la probabilidad inicial de que $\theta = \theta_i$ es p_i para $i = 1, ..., k$. Supóngase ahora, sin embargo, que antes de que el experimentador elija una acción a_j, puede observar los valores de una muestra aleatoria $X_1, ..., X_n$ seleccionada de una distribución que depende del parámetro θ.

Para $i = 1, ..., k$, se define $f_n(\mathbf{x} \mid \theta_i)$ como la función de densidad conjunta de las observaciones $X_1, ..., X_n$ cuando $\theta = \theta_i$. Después de haber observado el vector $\mathbf{x}$,

correspondiente a los valores en la muestra, la probabilidad a posteriori $p_i(\theta \mid \mathbf{x})$, será

$$p_i(\theta \mid \mathbf{x}) = P(\theta = \theta_i \mid \mathbf{x}) = \frac{p_i f_n(\mathbf{x} \mid \theta_i)}{\sum_1^k p_t f_n(\mathbf{x} \mid \theta_t)} \text{ para } i = 1, ..., k. \tag{5.2.3}$$

Por tanto, después de haber observado $\mathbf{x}$ de valores de la muestra, la recompensa $r(a_j, \mathbf{x})$ de seleccionar la acción a_j será

$$r(a_j, \mathbf{x}) = \sum_{i=1}^{k} p_i(\theta \mid \mathbf{x}) r_{ij} \text{ para } j = 1, ..., m. \tag{5.2.4}$$

Después de haber observado $\mathbf{x}$, se deduce que una decisión Bayes será una acción para la cual la recompensa de la ecuación 5.2.4 es un máximo. Dicha acción se denomina decisión Bayes respecto a la distribución a posteriori de θ.

En un problema de decisión múltiple de este tipo, una regla de decisión se define como una función $d(\mathbf{x})$ que especifica, para cada vector $\mathbf{x}$, una de las m acciones posibles $a_1, ..., a_m$.

Una regla de decisión d se denomina regla de decisión Bayes si, para cada vector posible $\mathbf{x}$, la decisión $d(\mathbf{x})$ es una decisión bayes respecto a la distribución a posteriori de θ. En otras palabras, cuando se utiliza una regla de decisión Bayes, la acción que se elige después de haber observado el vector $\mathbf{x}$ siempre es una acción para la cual la recompensa $r(a_j, \mathbf{x})$ es un máximo.

Antes de seleccionar las observaciones, la recompensa que el experimentador obtiene por utilizar una regla de decisión específica d se puede calcular como sigue: Para $j = 1, ..., m$, sea A_j el conjunto de todos los resultados $\mathbf{x}$ para los cuales $d(\mathbf{x}) = a_j$, esto es, para el que se elegirá la acción a_j. Por conveniencia, supóngase que las observaciones $X_1, ..., X_n$ tienen una distribución discreta y que $f_n(\mathbf{x} \mid \theta_i)$ representa su densidad conjunta cuando $\theta = \theta_i$. Si $f_n(\mathbf{x} \mid \theta_i)$ es realmente una función de densidad conjunta, entonces las sumas sobre valores de $\mathbf{x}$ que aparecen en el desarrollo dado aqui se deben remplazar por integrales.

Si $\theta = \theta_i$, el riesgo $r(d \mid \theta = \theta_i)$ de utilizar la regla d es

$$r(d \mid \theta = \theta_i) = \sum_{j=1}^{m} r_{ij} P[d(\mathbf{x}) = a_j \mid \theta = \theta_i] = \sum_{j=1}^{m} r_{ij} \sum_{x \in A_j} f_n(\mathbf{x} \mid \theta_i). \tag{5.2.5}$$

Puesto que la probabilidad a priori de que $\theta = \theta_i$ es p_i, la recompensa global $r(d)$ de utilizar la regla d será

$$r(d) = \sum_{i=1}^{k} p_i r(d \mid \theta = \theta_i) = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{x \in A_j} p_i r_{ij} f_n(\mathbf{x} \mid \theta_i). \tag{5.2.6}$$

Esta recompensa $r(d)$ es máxima cuando d es una regla de decisión Bayes.

Ejemplo 5.2.6 : *Determinación de una regla de decisión Bayes. Supóngase que en un gran cargamento de frutas, los únicos tres valores posibles de la proporción θ de piezas dañadas son* 0.1, 0.3 *y* 0.5 *y que hay tres acciones posibles a_1, a_2 y a_3. Supóngase, además, que sin pérdida de generalidad, tomamos pérdidas en lugar de recompensas, y encontrando el mínimo en lugar del máximo, las pérdidas de estas acciones son las siguientes:*

	a_1	a_2	a_3
$\theta = 0.1$	*0*	*1*	*3*
$\theta = 0.3$	*2*	*0*	*2*
$\theta = 0.5$	*3*	*1*	*0*

Supóngase, además, que teniendo en cuenta cargamentos anteriores del mismo distribuidor, se cree que las probabilidades a priori de los tres valores posibles de θ son las siguientes:

$$\begin{aligned} P(\theta = 0.1) &= 0.5 \\ P(\theta = 0.3) &= 0.3 \\ P(\theta = 0.5) &= 0.2 \end{aligned} \quad . \tag{5.2.7}$$

Finalmente, supóngase que se puede observar el número Y de piezas de frutas dañadas en una muestra aleatoria de 20 *piezas seleccionadas del cargamento. Se determinará una regla de decisión Bayes y se calculará la pérdida de esta regla. En la siguiente tabla se muestran las probabilidades a posteriori que se pueden tener después de observar la muestra:*

y	$P(\theta = 0.1 \mid Y = y)$	$P(\theta = 0.3 \mid Y = y)$	$P(\theta = 0.5 \mid Y = y)$
0	0.9961	0.0039	0
1	0.9850	0.0150	0
2	0.9444	0.0553	0.0002
3	0.8141	0.1840	0.0019
4	0.5285	0.4606	0.0109
5	0.2199	0.7393	0.0408
6	0.0640	0.8294	0.1066
7	0.0151	0.7575	0.2273
8	0.0031	0.5864	0.4105
9	0.0005	0.3795	0.6200
10	0.0001	0.2078	0.7921
11	0	0.1011	0.8989
12	0	0.046	0.954
13	0	0.020	0.980
14	0	0.009	0.991
15	0	0.004	0.996
16	0	0	1
17	0	0	1
18	0	0	1
19	0	0	1
20	0	0	1

Cuando $\theta = 0.1$, la distribución de Y es una distribución binomial con parámetros 20 y 0.1. La f.p. $g(y \mid \theta = 0.1)$ es la siguiente:

$$g(y \mid \theta = 0.1) = \begin{pmatrix} 20 \\ y \end{pmatrix} (0.1)^y (0.9)^{20-y}, \text{ para } y = 0, 1, ..., 20. \tag{5.2.8}$$

Cuando $\theta = 0.3$ o $\theta = 0.5$ la distribución de Y es una distribución binomial análoga y las expresiones para $g(y \mid \theta = 0.3)$ y $g(y \mid \theta = 0.5)$ tendrán una forma similar a la ecuación 5.2.8.

De la ecuación 5.2.3 se deduce que después de haber observado el valor de $Y = y$, la probabilidad a posteriori de que $\theta = 0.1$ será

$$P(\theta = 0.1 \mid Y = y) = \frac{(0.5)\, g(y \mid \theta = 0.1)}{(0.5)\, g(y \mid \theta = 0.1) + (0.3)\, g(y \mid \theta = 0.3) + (0.2)\, g(y \mid \theta = 0.5)}. \tag{5.2.9}$$

Se pueden escribir expresiones análogas para las probabilidades finales de que $\theta = 0.3$ y $\theta = 0.5$. Estas probabilidades a posteriori, para cada valor posible de y, se encuentran en la tabla.

Después de haber observado el valor de y, el riesgo $r(a_j, y)$ de cada acción posible a_j $(j = 1, 2, 3)$ se puede calcular aplicando la ecuación 5.2.4 y utilizando estas probabilidades a posteriori y la tabla de pérdidas dada al principio de este ejemplo. Por tanto, el riesgo $r(a_1, y)$ de elegir la acción a_1 será

$$r(a_1, y) = 2P(\theta = 0.3 \mid Y = y) + 3P(\theta = 0.5 \mid Y = y)$$

y	$r(a_1, y)$	$r(a_2, y)$	$r(a_3, y)$
0	0.0078	0.9961	2.9961
1	0.0300	0.9850	2.9850
2	0.1124	0.9446	2.9430
3	0.3737	0.8160	2.8103
4	0.9539	0.5394	2.5067
5	1.6010	0.2607	2.1383
6	1.9786	0.1706	1.8508
7	2.1969	0.2428	1.5603
8	2.4043	0.4136	1.1821
9	2.6190	0.6205	0.7605
10	2.7919	0.7922	0.4159
11	2.8989	0.8989	0.2022
12	2.954	0.954	0.092
13	2.980	0.980	0.040
14	2.991	0.991	0.018
15	2.996	0.996	0.008
16	3	1	0
17	3	1	0
18	3	1	0
19	3	1	0
20	3	1	0

el riesgo de elegir a_2 será

$$r(a_2, y) = P(\theta = 0.1 \mid Y = y) + P(\theta = 0.5 \mid Y = y)$$

y el riesgo de elegir a_3 será

$$r(a_3, y) = 3P(\theta = 0.1 \mid Y = y) + 2P(\theta = 0.3 \mid Y = y)$$

Los valores de estos riesgos para cada valor posible de y y cada acción posible se indican en la última tabla.

De los valores tabulados se pueden deducir las siguientes conclusiones: Si $y \leq 3$, entonces la decisión Bayes es a_1, si $4 \leq y \leq 9$, entonces la decisión Bayes es a_2 y si $y \geq 10$, entonces la decisión Bayes es a_3. En otras palabras, la regla de decisión Bayes d se define como sigue

$$d(y) = \begin{cases} a_1 \text{ si } y = 0,1,2,3, \\ a_2 \text{ si } y = 4,5,6,7,8,9, \\ a_3 \text{ si } y = 10,...,20. \end{cases} .$$

De la ecuación 5.2.6 resulta ahora que el riesgo $r(d)$ de utilizar la regla d es

$$\begin{aligned} r(d) &= \sum_{i=1}^{k} p_i r(d \mid \theta = \theta_i) = \sum_{i=1}^{k} \sum_{j=1}^{m} \sum_{y \in A_j} p_i r_{ij} f_n(\mathbf{y} \mid \theta_i) \\ &= (0.5) \sum_{y=4}^{9} g(y \mid \theta = 0.1) + (1.5) \sum_{y=10}^{20} g(y \mid \theta = 0.1) \\ &\quad + (0.6) \sum_{y=0}^{3} g(y \mid \theta = 0.3) + (0.6) \sum_{y=10}^{20} g(y \mid \theta = 0.3) \\ &\quad + (0.6) \sum_{y=0}^{3} g(y \mid \theta = 0.5) + (0.2) \sum_{y=4}^{9} g(y \mid \theta = 0.5) \\ &= 0.2423. \end{aligned}$$

Por tanto, el riesgo de la regla de decisión Bayes es $r(d) = 0.2423$.

Ejemplo 5.2.7 : *El valor de la información muestral. Supóngase ahora que fue necesario elegir una de las tres acciones a_1, a_2 o a_3 en el ejemplo anterior sin haber observado el número de piezas dañadas en una muestra aleatoria. En este caso, se encuentra a partir de la tabla de pérdidas r_{ij} y de las probabilidades iniciales 5.2.7 que los riesgos $r(a_1), r(a_2)$ y $r(a_3)$ de seleccionar cada una de las acciones a_1, a_2 y a_3 son los siguientes:*

$$\begin{aligned} r(a_1) &= 2(0.3) + 3(0.2) = 1.2, \\ r(a_2) &= (0.5) + (0.2) = 0.7, \\ r(a_3) &= 3(0.5) + 2(0.3) = 2.1. \end{aligned}$$

Por tanto, la decisión Bayes sin ninguna observación sería a_2 y el riesgo de la decisión sería 0.7.

Por el hecho de poder observar el número de piezas dañadas en una muestra aleatoria de 20 *piezas, se puede reducir que el riesgo de* 0.7 *a* 0.2423.

5.3 Determinación de una prima óptima en un modelo de decisión secuencial

Como se recordará de la exposición de este problema en el capítulo 1, vamos a considerar que la distribución de los reclamos es desconocida en el sentido de que se conoce la forma parámetrica de dicha distribución, pero no se conoce el valor concreto del parámetro que la determina. Para encontrar el valor de dicho parámetro, se parte de una conjetura inicial y se utiliza la información acerca de los reclamos ocurridos en cada período para modificar esta conjetura en cada época de decisión.

Empecemos reescribiendo los elementos del modelo.

5.3.1 El modelo

- $N(\pi)$: demanda de la póliza con prima π. $N(\pi)$ es decreciente.
- X_i : v.a. que toma el valor de 1 si el i-ésimo asegurado realiza un reclamo y 0 si no lo realiza. X_i se distribuye entonces como bernoulli con parámetro p,

$$X_i \sim Be(p),$$
$$\Pr[X_i = x_i] = p^{x_i}(1-p)^{1-x_i} \text{ con } x_i = 0,1; \text{ e } i = 1,..,N,$$

y $X = \sum_1^N X_i$ es la v.a. que se refiere al número de reclamos durante el período en que se aplica la prima π. X se distribuye como binomial con parámetros p y $N = N(\pi)$,

$$\begin{aligned} X &\sim b(N,p), \\ \Pr[X = x] &= \binom{N}{x} p^x (1-p)^{N-x} \text{ con } x = 0,1,...,N. \end{aligned}$$

Se supone que el período es suficientemente pequeño para que se dé a lo más un reclamo por cada asegurado durante ese período.

Así que $p = P$[ocurra un reclamo] es el parámetro desconocido en la distribución de los reclamos.

Si $\mathbf{x} = (x_1, x_2, ..., x_N)$ es una muestra de una distribución Bernoulli, su función de verosimilitud es

$$f(\mathbf{x} \mid p) = \prod_1^N p^{x_i}(1-p)^{1-x_i} = p^{\sum_1^N x_i}(1-p)^{N-\sum_1^N x_i} = p^x(1-p)^{N-x},$$

y la conjetura inicial del asegurador sobre este parámetro está representada por una distribución beta con parámetros r y n. Esto es debido a que los factores

que involucran al parámetro p asemejan a una distribución beta y una suposición válida, a partir de la función de verosimilitud de los reclamos ($f(\mathbf{x} \mid p)$), para una familia conjugada de las distribuciones a priori es la clase de las distribuciones beta.

$$\begin{aligned} p &\sim \mathbf{B}(r,n), \\ \Pr[p \mid r,n] &= \frac{\Gamma(r+n)}{\Gamma(r)\Gamma(n)} p^{r-1}(1-p)^{n-1}. \end{aligned}$$

(En adelante, el parámetro r representará el número de reclamos ocurridos en el período anterior y n el número de pólizas vendidas en ese período).

- Y_i : el monto del reclamo realizado por el $i-$ésimo asegurado. Suponemos que las Y_i son independientes y con distribución común exponencial con parámetro W

$$f_{Y_i}(y) = W\mathbf{e}^{-Wy}, \qquad y \geq 0,$$

Y es el monto total de los reclamos suponiendo que se efectuaron x reclamos.

$$Y = \sum_{i=1}^{x} Y_i,$$

Y se distribuye como una gama con parámetros x y W, es decir

$$Y \sim \mathbf{g}(x,W) = \frac{W}{\Gamma(x)}(yW)^{x-1}e^{-yW}.$$

Así que W es el parámetro desconocido de la distribución del monto de los reclamos y a partir de la función de verosimilitud,

$$f(\mathbf{y} \mid w) = \prod_{1}^{x} we^{-wy_i} = w^x e^{-w\sum_1^x y_i} = w^x e^{-wy},$$

vemos que al igual que con la distribución del número de reclamos, la conjetura inicial del asegurador acerca de este parámetro está representada por una distribución gama con parámetros α y β

$$W \sim \mathbf{g}(w \mid \alpha,\beta) = \frac{\beta}{\Gamma(\alpha)}(w\beta)^{\alpha-1}e^{-w\beta}.$$

Donde α es el número de reclamos recibidos en el período anterior y β el monto de los mismos.

5.3.2 Proceso de aprendizaje

El siguiente valor de p está dado por la distribución a posteriori

$$p(p \mid \mathbf{x}) = \frac{p(p) f(\mathbf{x} \mid p)}{m(\mathbf{x})} = \frac{h(\mathbf{x}, p)}{m(\mathbf{x})},$$

de acuerdo a la regla de Bayes .

Sin tomar en cuenta el valor de $m(\mathbf{x})$ y los factores que no involucran al parámetro p, se tiene que

$$h(\mathbf{x}, p) = p(p) f(\mathbf{x} \mid p) = \frac{\Gamma(r+n)}{\Gamma(r)\Gamma(n)} p^{x+r-1} (1-p)^{N+n-x-1} \propto p^{x+r-1} (1-p)^{N+n-x-1},$$

de donde la distribución a posteriori es una $Be(p \mid x+r, N+n-x)$ con $x = \sum_1^N x_i$. Puesto que la distribución a posteriori es una distribución beta, se sigue que la clase de las distribuciones beta es una familia conjugada para $\mathfrak{F}$, la clase de densidades binomial.

Análogamente, en el caso del monto de los reclamos,

$$h(\mathbf{y}, w) = p(w) f(\mathbf{y} \mid w) = \frac{\beta^{\alpha} w^{\alpha-1} e^{-w\beta}}{\Gamma(\alpha)} * w^{x} e^{-wy} I_{(0,\infty)}(w)$$

$$= \frac{\beta^{\alpha} w^{x+\alpha-1} e^{-w(y+\beta)}}{\Gamma(\alpha)} I_{(0,\infty)}(w).$$

Por lo que, fijándonos únicamente en los términos que involucran a w, el valor posterior de W está dado por una distribución $\mathbf{g}(w \mid \alpha + x, \beta + y)$, ya que como deciamos antes, la función de verosimilitud para las densidades exponenciales, se asemeja a una densidad gama suponiendo que y es el monto de los x reclamos ocurridos.

5.3.3 Determinación de la prima óptima

Suponemos que el objetivo del asegurador es maximizar su utilidad descontada esperada sobre el horizonte de planeación T.

Considerando el proceso de aprendizaje descrito antes, el estado (x, y) se transforma en (α, β, r, n) pues la distribución del número de reclamos x depende finalmente de los parámetros r y n, y la distribución del monto de los reclamos y depende de α y β.

En un solo período t, la utilidad esperada cuando se aplica la prima π está dada por

$$\Psi(\pi \mid \alpha, \beta, r, n) = \pi N(\pi) - \Psi_1(\pi \mid \alpha, \beta, r, n),$$

donde $\Psi_1(\pi \mid \alpha, \beta, r, n) = E[Y]$. Es decir, la utilidad es el ingreso recibido por la venta de $N(\pi)$ pólizas, a una prima π, menos el valor esperado del monto de los siniestros que se deben cubrir. Este valor esperado se puede obtener a través de

$$E[Y] = E[E[Y \mid X]],$$

y para obtener la esperanza condicional $E[Y \mid X]$ requerimos integrar la densidad condicional de $Y \mid X$, multiplicada por y, sobre todos los posibles valores de $y \in [0, \infty)$. Así, obtenemos:

$$\begin{aligned} E[Y \mid X] &= \int_0^\infty \left[\int_0^\infty y\mathbf{g}(y \mid x, w)\, dy \right] \mathbf{g}(w \mid \alpha, \beta)\, dw \\ &= \int_0^\infty \int_0^\infty y\mathbf{g}(y \mid x, w)\, \mathbf{g}(w \mid \alpha, \beta)\, dy\, dw. \end{aligned}$$

Ahora, para obtener el valor esperado de Y necesitamos integrar esta esperanza condicional sobre todos los posibles valores de X, es decir

$$\Psi_1(\pi \mid \alpha, \beta, r, n) =$$
$$\int_0^1 \int_0^\infty \int_0^\infty \sum_{x=0}^{N(\pi)} y\mathbf{g}(y \mid x, w)\, \mathbf{g}(w \mid \alpha, \beta)\, \mathbf{b}(x \mid N(\pi), p)\, \mathbf{B}(p \mid r, n)\, dy\, dw\, dp.$$

Si no se tomaran en cuenta los valores ocurridos de X y Y en un proceso de aprendizaje, entonces la prima π^0 se aplicaría para todo t. π^0 podría obtenerse maximizando la utilidad $\Psi(\pi \mid \cdot)$. Suponiendo cierta regularidad en $N(\pi)$, que $\pi^0(\cdot) > 0$, y que $\pi N(\pi)$ está acotada inferiormente, sabemos que existe π^0 que maximiza $\Psi(\alpha, \beta, n, r)$. Definamos

$$Q(\alpha, \beta, r, n) = \Psi\left(\pi^0 \mid \alpha, \beta, r, n\right).$$

Como tiene lugar un proceso de aprendizaje, el asegurador debe considerar no sólo el efecto de la prima que determine en la utilidad inmediata, sino su efecto en los parámetros futuros. Si en algún período t el número de pólizas vendido es $N(\pi)$, el número de reclamos es x, y el valor de estos reclamos es y, entonces los parámetros de las distribuciones al inicio del período $t+1$ son $\alpha + x$, $\beta + y, r + x$, $n + N(\pi) - x$.

Sea $V_t^\lambda(\alpha, \beta, r, n)$ la utilidad esperada descontada del tiempo t en adelante dado que los parámetros son (α, β, r, n) en ese período. Por el algoritmo de programación dinámica sabemos que

$$\begin{aligned} & V_t^\lambda(\alpha, \beta, r, n) \\ = \ & \max\left\{ \Psi(\pi \mid \alpha, \beta, r, n) + \lambda E\left[V_{t+1}^\lambda(\alpha + x, \beta + y, r + x, n + N(\pi) - x) \right] \right\}, \end{aligned}$$

para $t = 1, 2, \ldots T-1$, y

$$V_T^\lambda(\alpha, \beta, r, n) = Q(\alpha, \beta, r, n).$$

Considerando las ditribuciones que hemos venido utilizando, tenemos que

$$\begin{aligned} & E\left[V_{t+1}^{\lambda}\left(\alpha+x,\beta+y,r+x,n+N\left(\pi\right)\right)\right] \\ = & \int_0^1\int_0^{\infty}\int_0^{\infty}\sum_{x=0}^{N(\pi)} V_{t+1}^{\lambda}\left(\alpha+x,\beta+y,r+x,n+N\left(\pi\right)-x\right) \\ & \cdot\mathbf{g}\left(y\mid x,w\right)\mathbf{g}\left(w\mid\alpha,\beta\right)\mathbf{b}\left(x\mid N\left(\pi\right),p\right)\mathbf{B}\left(p\mid r,n\right)dy\,dw\,dp. \end{aligned}$$

Aplicando el algoritmo de programación dinámica, obtendremos entonces la prima óptima para cada período $\pi_t^*\left(\alpha,\beta,r,n\right)$ como un valor maximizador en cada paso.

5.3.4 Efectos del proceso de aprendizaje en las primas óptimas

Aquí se trata de demostrar que $\pi_t^*\left(\cdot\right)\leq\pi^0\left(\cdot\right)$ para toda t. Intuitivamente, esta desigualdad quiere decir que mientras las primas tengan un precio menor, estas podrán venderse más, lo que implica obtener mayor información sobre los parámetros desconocidos. Así, el proceso de aprendizaje brinda una ventaja adicional sobre el proceso que fija una prima igual para todos los períodos.

Con base en lo anterior, nuestro objetivo será el de demostrar el siguiente teorema, que requiere de la demostración de dos lemas. El primer lema demuestra las condiciones bajo las cuales se cumplen las hipótesis del teorema y el segundo demuestra que nuestro modelo las cumple.

Teorema 5.3.1 *Sea G_t una función decreciente de π, dada por*

$$G_t\left(\pi\mid\alpha,\beta,r,n\right)=\lambda E\left[V_{t+1}^{\lambda}\left(\alpha+x,\beta+y,r+x,n+N\left(\pi\right)-x\right)\right],$$

para $t=1,2,\ldots,T-1$, $0\leq\lambda\leq1$ y

$$G_T\left(\pi\mid\alpha,\beta,r,n\right)=0,$$

entonces $\pi^0\left(\alpha,\beta,r,n\right)\geq\pi_t^*\left(\alpha,\beta,r,n\right)$ para toda $t=1,...,T-1$.

Demostración. Definamos la función $F_t\left(\pi,\lambda\right)$ como $F_t\left(\pi,\lambda\right)=\Psi\left(\pi\mid\cdot\right)+\lambda G_t\left(\pi\mid\cdot\right)$, y supongamos que $\pi\left(\lambda\right)$ maximiza $F_t\left(\pi,\lambda\right)$. Entonces, en $\pi\left(\lambda\right)$ tenemos

$$\frac{\partial}{\partial\pi}\Psi\left(\pi\mid\cdot\right)+\lambda\frac{\partial}{\partial\pi}G_t\left(\pi\mid\cdot\right)=0$$

$$\frac{\partial^2}{\partial\pi^2}\Psi\left(\pi\mid\cdot\right)+\lambda\frac{\partial^2}{\partial\pi^2}G_t\left(\pi\mid\cdot\right)<0.$$

Ahora bien, como

$$F_t\left(\pi,0\right)=\Psi\left(\pi\mid\cdot\right)$$

y

$$F_t(\pi, 1) = \Psi(\pi \mid \cdot) + G_t(\pi \mid \cdot),$$

se tiene que el valor máximo de π en $\lambda = 1$, corresponde al valor de π_t^*, es decir, $\tilde{\pi}(1) = \pi_t^*$ y de la misma manera, en $\lambda = 0$, corresponde al valor de π^0, es decir, $\tilde{\pi}(0) = \pi^0$. Así, la prueba consiste solo de demostrar que la función $\tilde{\pi}(\lambda)$ es decreciente, es decir, que $\frac{d}{d\lambda}\tilde{\pi}(\lambda)$. Para esto notamos que

$$\operatorname{sign}\left(\frac{d}{d\lambda}\tilde{\pi}(\lambda)\right) = \operatorname{sign}\left(\frac{\partial^2}{\partial\pi\lambda}F_t(\pi,\lambda)\right) = \operatorname{sign}\left(\frac{\partial}{\partial\pi}G_t(\pi \mid \cdot)\right) \leq 0.$$

■

.

En el teorema anterior se dió por supuesto que la función $G(\pi \mid \cdot)$ es no creciente. En este trabajo usaremos el siguiente lema cuya demostración se puede encontrar en Blackwell (1951) [5], (1953) [6], Marschak y Miyasawa (1968) [7]. Este lema establece condiciones bajo las cuales $G(\pi \mid \cdot)$ es una función no creciente de π. Lo que se demuestra entonces es que si $N(\pi_1)$ y $N(\pi_2)$ son enteros que satisfacen $N(\pi_1) > N(\pi_2)$, entonces $G(\pi_1) \geq G(\pi_2)$. Claramente, si $\pi_2 > \pi_1$, entonces $N(\pi_1) > N(\pi_2)$, donde asumimos que $N(\pi_1)$ y $N(\pi_2)$ son enteros. De aqui en adelante, $N(\pi_1) = N_1$ y $N(\pi_2) = N_2$.

Lema 5.3.2 *Si existe una función $\eta(x', y' \mid x, y)$ que no dependa de p o W, y que satisfaga*

$$f(x', y' \mid p, W, N_2) = \int_0^\infty \sum_{x=0}^{N_1} \eta(x', y' \mid x, y) f(x, y \mid p, W, N_1)\, dy \tag{5.3.10}$$

y

$$\int_0^\infty \sum_{x'=0}^{N_2} \eta(x', y' \mid x, y)\, dy' = 1, \tag{5.3.11}$$

entonces el experimento $f(x, y \mid \cdot, N_1)$ es más informativo que $f(x', y' \mid \cdot, N_2)$ en el sentido de que $G(\pi_1 \mid \cdot) \geq G(\pi_2 \mid \cdot)$.

La idea del lema anterior es que la determinación de la prima π realmente define un experimento que consiste en una muestra aleatoria de la distribución $f[x, y \mid p, W, N(\pi)] = bin(x \mid p, N(\pi))\, gama(y \mid x, W)$. Asi que, cuando ponderamos π_1 contra π_2 (asumiendo, sin pérdida de generalidad, que $\pi_1 < \pi_2$), realmente ponderamos, el experimento $f[x, y \mid p, W, N(\pi_1)]$ contra el experimento $f[x', y' \mid p, W, N(\pi_2)]$, donde x y y (x' y y') son el número de reclamos y el valor monetario del total de los reclamos dado que $N(\pi_1)$ y $N(\pi_2)$ polizas de seguros se vendieron, respectivamente. Ahora vamos a demostrar que nuestro modelo satisface las hipótesis del lema anterior.

Lema 5.3.3 *La función $\eta(x', y' \mid x, y)$ definida en , satisface las condiciones 5.3.10 y 5.3.11 , si $N_1 > N_2$.*

Demostración. Construimos la función $\eta(\cdot \mid \cdot)$ como sigue. Primero definimos las funciones $\eta_1(x' \mid x)$ y $\eta_2(y' \mid x', x, y)$ como

$$\eta_1(x' \mid x) = \begin{cases} \frac{\binom{N_2}{x'}\binom{N_1-N_2}{x-x'}}{\binom{N_1}{x}} \text{ si } N_2 - N_1 - x \leq x' \leq x \\ 0 \text{ en otro caso.} \end{cases}$$

y

$$\eta_2(y' \mid x', x, y) = \begin{cases} \frac{\Gamma(x)}{\Gamma(x')\Gamma(x-x')} \left(\frac{y'}{y}\right)^{x'-1} \left(1 - \frac{y'}{y}\right)^{x-x'-1} \cdot \frac{1}{y} \text{ si } 0 \leq y' \leq y, 0 \leq x' \leq x \\ 0 \text{ en otro caso.} \end{cases}$$

Después definimos

$$\eta_1(x', y' \mid x, y) = \eta_1(x' \mid x)\, \eta_2(y' \mid x', x, y).$$

Para ver esto último, sea Y' una v.a.i. que se distribuye $gama(\cdot \mid x', W)$, Z' una v.a.i. que se distribuye $gama(\cdot \mid x - x', W)$ y sea $Y = Y' + Z'$. Hay que encontrar $f_{Y'|Y}(y' \mid y)$.

Como $f_{Y'|Y}(y' \mid y) = \frac{f_{Y',Y}(y',y)}{f_Y(y)}$, hay que encontrar las dos funciones de la división. Primero vemos que

$$f_{Y',Y}(y', y) = /J/ f_{Y',Z'}(y', y - y') = f_{Y'}(y')\, f_{Z'}(y - y')$$

$$= \frac{w}{\Gamma(x')} (wy')^{x'-1} e^{-wy'} \cdot \frac{w}{\Gamma(x - x')} (w(y - y'))^{x-x'-1} e^{-w(y-y')}$$

$$= \frac{w^x}{\Gamma(x')\Gamma(x - x')} e^{-wy} y'^{(x'-1)} (y - y')^{x-x'-1},$$

luego,

$$f_Y(y) = \int_0^\infty f_{Y,Y'}(y, y')\, dy' = \frac{w^x e^{-wy}}{\Gamma(x')\Gamma(x - x')} \int_0^\infty y'^{(x'-1)} (y - y')^{x-x'-1}\, dy'$$

$$= \frac{y^{x-2} w^x e^{-wy}}{\Gamma(x')\Gamma(x - x')} \int_0^\infty \left(\frac{y'}{y}\right)^{x'-1} \left(1 - \frac{y'}{y}\right)^{x-x'-1} dy'.$$

Sea $u = \frac{y'}{y}$; $yu = y'$; diferenciando respecto a u y y', $ydu = dy\prime$; $du = \frac{dy'}{y}$ y como $0 < y' \leq y < \infty$; entonces $0 < \frac{y'}{y} \leq 1$.

$$f_Y(y) = \frac{y^{x-1} w^x e^{-wy}}{\Gamma(x)} \int_0^1 \frac{\Gamma(x)}{\Gamma(x')\Gamma(x - x')} u^{x'-1} (1 - u)^{x-x'-1}\, du$$

$$= \frac{w}{\Gamma(x)} (wy)^{x-1} e^{-wy} = gama(y \mid x, w) \ .$$

Por lo que finalmente,

$$f_{Y'|Y}(y' \mid y) = \frac{f_{Y',Y}(y', y)}{f_Y(y)} = \frac{\Gamma(x)}{\Gamma(x')\Gamma(x-x')} \left(\frac{y'}{y}\right)^{x'-1} \left(1 - \frac{y'}{y}\right)^{x-x'-1} \cdot \frac{1}{y},$$

que es una función $Beta(u \mid x', x-x')$ y $u = \frac{y'}{y}$.

Análogamente, sea X' una v.a.i. que se distribuye $bin(\cdot \mid N_2, p)$, Z una v.a.i. que se distribuye $bin(\cdot \mid N_1 - N_2, p)$ y sea $X = Z + X'$. Hay que encontrar $f_{X'|X}(x' \mid x)$.

Primero vemos que

$$\begin{aligned} f_{X',X}(x', x) &= f_{X',Z'}(x', x-x') = f_{X'}(x') f_{Z'}(x-x') \\ &= \binom{N_2}{x'} \binom{N_1 - N_2}{x - x'} p^{x'} (1-p)^{N_2 - x'} p^{x-x'} (1-p)^{N_1 - N_2 - x + x'} \\ &= \binom{N_2}{x'} \binom{N_1 - N_2}{x - x'} p^{x} (1-p)^{N_1 - x}; \end{aligned}$$

luego,

$$f_X(x) = \sum_{x'=0}^{x} \binom{N_2}{x'} \binom{N_1 - N_2}{x - x'} p^x (1-p)^{N_1 - x} = p^x (1-p)^{N_1 - x} \sum_{x'=0}^{x} \binom{N_2}{x'} \binom{N_1 - N_2}{x - x'},$$

y como en general se cumple

$$\sum_{j=0}^{n} \binom{a}{j} \binom{b}{n-j} = \binom{a+b}{n},$$

tenemos,

$$f_X(x) = \binom{N_1}{x} p^x (1-p)^{N_1 - x} \ .$$

Por lo que finalmente,

$$f_{X'|X}(x' \mid x) = \frac{f_{X',X}(x', x)}{f_X(x)} = \frac{\binom{N_2}{x'} \binom{N_1 - N_2}{x - x'}}{\binom{N_1}{x}},$$

que es una función $Hipergeométrica(x \mid N_1, N_2, x')$.

Por último, como

$$\begin{aligned} \eta_1(x' \mid x) &= f_{X'|X}(x' \mid x), \ \eta_2(y' \mid x', x, y) = f_{Y'|Y}(y' \mid y) \text{ y } f(x, y \mid p, W, N_1) \\ &= f_X(x) f_Y(y), \end{aligned}$$

vemos que la sustitución de estos términos en la relación 5.3.10 nos da

$$f\left(x', y' \mid p, W, N_2\right) = \int_0^{\infty} \sum_{x=0}^{N_1} f_{X'|X}\left(x' \mid x\right) f_{Y'|Y}\left(y' \mid y\right) f_X\left(x\right) f_Y\left(y\right) dy,$$

$$\int_0^{\infty} \sum_{x=0}^{N_1} f_{X',X}\left(x', x\right) f_{Y',Y}\left(y', y\right) dy = f_{X'}\left(x'\right) f_{Y'}\left(y'\right),$$

y de la relación 5.3.11 claramente vemos que se cumple porque $\eta_2\left(y' \mid x', x, y\right)$ y $\eta_1\left(x' \mid x\right)$ son funciones de densidad y por consiguiente al sumarlas e integrarlas sobre todos sus valores, nos da 1. ■

Bibliografía

[1] M.L. Puterman (1994), *Markov Decision Processes: Discrete stochastic dynamic programming*, Wiley-Interscience Pub., New York.

[2] I. Venezia y H. Levy (1983), Optimal multi-period insurance contracts, *Insurence Math Econ.* **2,** 199-208.

[3] I. Venezia, Optimal insurance rates when the distribution of claims is unknown, *J. Applied Probability* **16,** 678-684.

[4] M.H. Degroot (1988), *Probabilidad y estadística,* Addison-Wesley Iberoamericana, Wilmington, Delaware.

[5] D. Blackwell (1951), Comparison of experiments. *Proc. 2nd Berkeley Symp. Math. Statis. Prob.*, 93-102.

[6] D. Blackwell (1953), Equivalent comparisons of experiments, *Ann. Math. Stat.* **24,** 265-273.

[7] J. Marschak y K. Miyasawa (1968), Economic comparability of economic systems, *Internat. Econom. Rev.* **9,** 137-174.