Disponible en www.sciencedirect.com

www.cya.unam.mx/index.php/cya

Contaduría y Administración 61 (2016) 176–201

www.contaduriayadministracionunam.mx/

Cálculo del valor en riesgo operacional mediante redes bayesianas para una empresa financiera

Operational value at risk by bayesian networks for a financial firm

Griselda Dávila Aragón, Francisco Ortiz Arango *
y Fernando Cruz Aranda

Universidad Panamericana, Méxcio

Recibido el 11 de noviembre de 2014; aceptado el 3 de enero de 2015
Disponible en Internet el 24 de octubre de 2015

Resumen

El objetivo del presente trabajo es plantear la metodología basada en el uso de redes bayesianas (RB) para identificar y cuantificar los factores de riesgo operacional (RO) asociados al proceso de transacciones financieras a través de medios electrónicos en una empresa financiera. El modelo de RB desarrollado se ejemplifica con datos de eventos simulados en un periodo equivalente a 6 años a partir de información proporcionada por expertos en este tipo de procesos. Lo anterior representa una de las principales ventajas del uso de RB, pues permite modelar las relaciones causa-efecto entre los diferentes factores de RO. Finalmente se realiza el cálculo del valor en riesgo operacional *(OpVaR)* para el ejemplo, en el que se incorporan factores de interacción que no son considerados en el modelo tradicional, proporcionando mejores condiciones de credibilidad a este valor.

Palabras clave: Riesgo operacional; Redes bayesianas; Transacciones electrónicas; Árboles de derivación

Abstract

The aim of this paper is to outline the methodology based on the use of Bayesian networks (BN) to identify and quantify operational risk (OR) factors associated with processing financial transactions through

* Autor para correspondencia.
Correo electrónico: fortizar@up.edu.mx (F. Ortiz Arango) .

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

http://dx.doi.org/10.1016/j.cya.2015.09.009

electronic means in a financial company. BN model developed is exemplified with data from simulated events equivalent to six years period, from information provided by experts in this type of process. This represents one of the main advantages of using BR, they allow modeling the cause-effect relationships between different OR factors. Finally operational value at risk (OpVaR) for the example is calculated, where interacting factors that are not considered in the traditional model are incorporated, providing better conditions of credibility to this value.

Keywords: Operational risk; Bayesian networks; Electronic transactions; Derivation trees

Introducción

Las instituciones financieras, al igual que muchas otras empresas en diferentes sectores, han tomado conciencia a través de los años de los riesgos e incertidumbres que surgen de las fallas en sus operaciones y recientemente, de manera particular, de las fallas en los sistemas de información y la infraestructura tecnológica, paradigmas de suma importancia en la actualidad. De tal manera que los fraudes, la interrupción de la actividad operativa y la responsabilidad legal se han convertido en una amenaza constante para cualquier empresa. El riesgo operacional (RO) es inherente a toda actividad en que intervengan personas, procesos y plataformas tecnológicas, y es el único riesgo que no surge de la toma de una posición de incertidumbre financiera.

Considerando que los procesos y sistemas son desarrollados y administrados por personas, son ellas quienes causan los eventos de RO al hacer algo que no deberían haber hecho, o no hacer algo que debieron hacer. La intervención de las personas en los procesos de operación es una de las razones por las que la cuantificación del RO es tan complicada.

Otro motivo es la falta de información estadísticamente confiable, dado el corto período de tiempo de datos históricos de pérdidas asociadas al RO, el rol del ambiente de control interno, el cual naturalmente cambia, y el importante papel de los eventos de poca frecuencia pero alta severidad.

La limitante de encontrar bases de datos completas que contemplen los diferentes aspectos del RO y una compleja interacción entre las variables de riesgo que lo caracterizan hace que la medición de este riesgo requiera de técnicas dinámicas de medición.

Bajo estas circunstancias, el enfoque bayesiano es una alternativa viable para el análisis de riesgos en condiciones de información insuficiente. Los modelos bayesianos son modelos causa-efecto dinámicos que incorporan información inicial a través de una distribución de probabilidad *a priori*, mediante la cual se puede incluir información subjetiva en la toma de decisiones, como la opinión de expertos, el juicio de analistas o las creencias de especialistas. Esta información permite la actualización del modelo incorporando la información más reciente.

En este trabajo se propone un modelo de medición avanzada de RO, basado en el uso de redes bayesianas (RB), para cuantificar el RO en uno de los principales procesos de una empresa financiera donde anualmente se registran millones de transacciones electrónicas. El modelo propuesto

se alimenta con datos obtenidos de entrevistas con expertos[1] en los procesos que se modelan.En la bibliografía disponible existen trabajos, como los de Reimer y Neu (2003), Leippold (2003), Neil, Marquez y Fenton (2004) y Alexander (2002), que abordan de manera general la aplicación de las redes bayesiana a la administración del RO. Sin embargo, no se especifica cómo clasificar los eventos de riesgo, cómo identificarlos, cómo cuantificarlos ni cómo calcular el capital económico de manera consistente. En este trabajo se establecen las estructuras de información sobre eventos de RO de manera que sea posible identificar, cuantificar y medir el RO, cambiando el supuesto de independencia de eventos de RO, para modelar de manera realista el comportamiento causal de los eventos asociados al RO. Para lograr esto fue necesario estudiar la correlación entre factores de riesgo a fin de desarrollar un modelo de RB que permitiera identificar y cuantificar el RO del proceso de transacciones electrónicas en la empresa.

El presente trabajo se organiza de la siguiente manera: en la segunda sección se hace el planteamiento del proceso de construcción de una RB, sus características y las bondades del uso de estas. En la tercera sección se expone la problemática que se pretende resolver y se construye la RB a partir del análisis de los factores de riesgo asociados al proceso en cuestión; dada la complejidad del proceso se construyen 2 RB: una para la frecuencia y otra para la severidad. En la cuarta sección se lleva a cabo la cuantificación de cada nodo de las redes y se obtienen las probabilidades *a priori* a partir de la opinión de expertos; este es el único medio posible para obtener las probabilidades correspondientes, dada la falta de información estadística; a partir de esta primera información se calculan las probabilidades *a posteriori* a través de algoritmos de inferencia bayesiana, utilizando árboles de derivación; el objetivo final es determinar el cálculo del capital de RO que debe considerar la empresa, para lo cual se analizan distintos escenarios. Finalmente se presentan las conclusiones.

Redes bayesianas

En teoría de probabilidad clásica se asume que las estadísticas de la muestra pertenecen a cierta población con una distribución específica, la cual es definida por el conjunto de parámetros con un valor fijo. La tarea para el estadístico es estimar los parámetros lo mejor posible basándose en los datos disponibles, y cuando es posible, realizar experimentos varias veces y obtener así una muestra suficientemente grande para asignar valores a estos parámetros.

Thomas Bayes (1702-1761) estudió e investigó el problema de la determinación de la probabilidad de las causas a través de los efectos que se observan.

Los estadísticos bayesianos permiten a los parámetros ser variables aleatorias. Las afirmaciones hechas sobre las características de una población son necesariamente dependientes no solo de las observaciones empíricas o de los datos (información objetiva), sino también de cualquier conocimiento disponible para el estadístico previo al inicio de las observaciones (información subjetiva). Este conocimiento puede presentarse en forma de datos de un lugar diferente y se considera que tiene un cierto grado de relevancia para la población observada. También puede provenir de información obtenida de partes interesadas y expertos, es decir, cuya familiaridad con el tema los hace una fuente creíble. Como señala Aczel-Sounderpandian (2009), el enfoque bayesiano permite al estadístico complementar la información obtenida de muestreo con información previa obtenida de un particular (experto); el enlace matemático entre las probabilidades

[1] Colaboradores de empresas financieras que tienen la experiencia e información sobre la operación y administración de las líneas de negocio vinculadas con el proceso analizado.

asociadas a los datos observados y las probabilidades asociadas a la información obtenida de los expertos se logra a través del teorema de Bayes.

Allí donde la probabilidad clásica trata principalmente con la evaluación de declaraciones incondicionales de probabilidad, como: «la probabilidad del evento A es x» denotada por $P(A) = x$, el vocabulario bayesiano se expande a utilizar probabilidades condicionales en sus afirmaciones, como: «la probabilidad del evento A dado que el evento B ha ocurrido es y», y se denota por $P(A|B) = y$.

La manipulación de tales probabilidades consiste en tratarlas como funciones de variables que usan las reglas del cálculo de probabilidades. La regla fundamental del cálculo de probabilidades y piedra angular en la estadística bayesiana es el teorema de probabilidad condicional de los eventos A y B:

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

Dado que la función $P(A \cap B)$ es simétrica, el teorema se puede expresar también de la siguiente manera:

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

Igualando ambas ecuaciones y despejando $P(B|A)$, se obtiene el teorema de Bayes: $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

El teorema anterior es interpretado como: la probabilidad *a posteriori* $P(B|A)$ es igual a la probabilidad *a priori* $P(B)$ multiplicada por la razón $\frac{P(A|B)}{P(A)}$; es decir, la información previa acerca de B puede ser utilizada para revisar la probabilidad de B.

Al aplicar el teorema de Bayes a las distribuciones de variables aleatorias de un modelo, por ejemplo, sean $X = A$, $\Theta = B$ variables aleatorias y es tal que $x \in X$ y $\theta = \Theta$ tomadas de la muestra de la población con una distribución de probabilidad de Θ, entonces la probabilidad condicional estará dada por:

$$P(\theta \in \Theta | x \in X) = \frac{P(x \in X | \theta \in \Theta)\, P(\theta \in \Theta)}{P(x \in X)}$$

El teorema de Bayes en términos de la información objetiva y subjetiva. $P(\theta \in \Theta)$ es la información *a priori* de la población y es subjetiva. En este caso es una probabilidad incondicional que representa la incertidumbre acerca de $\theta \in \Theta$. La función $P(x \in X|\theta \in \Theta)$ es comúnmente llamada la verosimilitud del conjunto de valores en x denotada también como $L(x \in X|\theta \in \Theta)$ y se interpreta como la probabilidad de observar un conjunto de ciertos datos x dado que ciertas características de la población $\theta \in \Theta$ son ciertas.

Al combinar la información subjetiva con las observaciones empíricas (la información *a priori* y las verosimilitudes), se obtiene la probabilidad *a posteriori*, es decir, la probabilidad de $\theta \in \Theta$ tome ciertos valores dado que las observaciones de x han ocurrido, lo cual se denota $P(\theta \in \Theta | x \in X)$. Si se suma $P(x \in X|\theta \in \Theta)\, P(\theta \in \Theta)$, se obtiene la probabilidad marginal de $x \in X$:

$$P(x \in X) = \sum_{i=1}^{n} P(x_i \in X|\theta \in \Theta)\, P(\theta \in \Theta)$$

La probabilidad incondicional de $P(x \in X)$ sirve como una constante de escala. La regla puede ser reexpresada de la siguiente manera:

$$P(x \in X, \theta \in \Theta) = P(x \in X|\theta \in \Theta)\, P(\theta \in \Theta) = L(\theta \in \Theta|x \in X)\, P(x \in X)$$

La distribución conjunta de los datos observados y los parámetros es igual a la densidad de las observaciones dados los parámetros por la densidad de los parámetros que equivale a la distribución a posteriori de los parámetros dados los datos observados por la densidad marginal de los datos. $L(\theta \in \Theta | x \in X)$ es la función de distribución a posteriori, es decir la probabilidad condicional de $\theta \in \Theta$ dado que ha ocurrido $x \in X$.

De esta manera, la distribución a posteriori puede ser expresada como:

$$P(\theta \in \Theta | x \in X) \, \alpha \, L(x \in X | \theta \in \Theta) \, P(\theta \in \Theta)$$

La distribución a posteriori es equivalente al producto de la función de verosimilitud $x \in X$ de los datos observados, dados $\theta \in \Theta$ por la probabilidad de $\theta \in \Theta$, el conocimiento *a priori*.

Intuitivamente la idea bayesiana indica que las conjeturas acerca de la población son una combinación del conocimiento previo de la población y las observaciones hechas con respecto a esa misma población.

No es solo la información *a priori* la que tiene influencia; el grado de confianza en la información del experto también tiene un efecto. El análisis bayesiano simplemente formaliza este concepto: las densidades posteriores toman en cuenta en menor o mayor grado la información de las creencias dependiendo de la confianza que sobre esta se tenga.

No hay una respuesta simple a la pregunta sobre cuánta información de la muestra debe ser usada y qué tanta confianza se debe tener en las creencias previas. La única manera en que se puede saber cuáles son las mejores elecciones para el problema en análisis, como en cualquier modelo estadístico, es llevar a cabo un análisis de «*back testing*».

El teorema de Bayes ha sido usado durante muchos años en numerosas aplicaciones, entre ellas cálculos de las primas de seguros (Hossack et al., 1999). Es una forma racional de revisar las creencias a la luz de la evidencia observada. La subjetividad en sus cálculos es cuestionada en muchas ocasiones.

Una vez que se obtiene información *a priori* relevante sobre las posibles causas y probabilidades condicionales asociadas a cada efecto, el modelo es llamado «red bayesiana».

Las redes de creencias, las redes causales así como los modelos generativos son tipos de grafos dirigidos. Los grafos dirigidos son una combinación de la teoría de la probabilidad y la teoría de grafos. Es el resultado de la convergencia de la evolución de la modelación estadística, ingeniería e inteligencia artificial que inició en la década de los ochenta (Alexander, 2002).

Una RB es un grafo acíclico dirigido; en una RB, los nodos representan las variables de interés y las aristas son los enlaces causales o de influencia entre las variables. Asociado con cada nodo se tiene una tabla de probabilidad del nodo, una distribución estadística o una función parametrizada (Cardozo, 2011). En el caso de una tabla de probabilidad del nodo, la relación se rige por un conjunto de valores de probabilidad condicionales que modelan la relación incierta entre el nodo y sus nodos padre junto con cualquier incertidumbre presente en esa relación.

Inicialmente los cálculos excesivos basados en la teoría de probabilidad hacían inviable su uso; sin embargo, la utilización de independencia condicional en la teoría de grafos y el reciente desarrollo en algoritmos eficientes para la propagación de evidencia a través de estructuras gráficas han hecho que este campo sea mucho más factible computacionalmente.

Las RB han sido estudiadas como una herramienta potencial para varias aplicaciones en la gestión de riesgos. Sus características, que permiten la combinación de la opinión subjetiva de expertos, de datos observados y los modelos de causa y efecto, las hacen especialmente adecuadas para la investigación y captura del funcionamiento de las instituciones financieras. Aunque su uso hasta ahora ha sido limitado a áreas específicas, su aplicación a los riesgos empresariales más amplios está siendo cada vez más documentada, especialmente en el área de RO. La teoría detrás

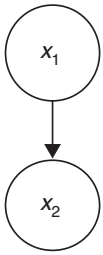

Figura 1. Grafos dirigidos.

de las RB combina la teoría de probabilidad bayesiana y la noción de independencia condicional para representar dependencias entre las variables.

La clave para un diseño exitoso de una RB es la descomposición significativa de un dominio del problema en un conjunto de proposiciones causales o condicionales sobre el dominio. En lugar de llevar a cabo la laboriosa y difícil tarea de calcular la distribución de probabilidad conjunta de todas las variables de interés, se aplica el principio «divide y vencerás» y se toman las especificaciones parciales del modelo que son en sí significativas en el dominio del experto.

Posteriormente se modela la tabla de probabilidad del nodo para cada variable, lo cual puede ser hecho usando datos históricos o solamente pidiéndole a un experto que proporcione una serie de estimaciones subjetivas, las cuales estarán basadas en su conocimiento y experiencia.

Una vez construida la RB, esta puede ser «ejecutada» usando un algoritmo apropiado de propagación. Cuando la RB se ejecuta, el efecto de los datos ingresados dentro de uno o más nodos son propagados por toda la red, en todas las direcciones y la distribución marginal de los nodos es actualizada. Esto hace al modelo ideal para un análisis de escenarios.

La definición matemática de RB está dada por los siguientes elementos:

a. Un conjunto de variables aleatorias conectadas por un conjunto de arcos (modelo de grafos).La figura 1 ejemplifica el uso de modelos en forma de grafos dirigidos. En ella, $x_1 = A$ y $x_2 = B$ son los nodos y representan las variables aleatorias x_1 y x_2. El segmento dirigido de x_1 a x_2 implica una relación de causalidad entre la variable aleatoria x_1 y x_2 e indica que un cambio en lo que se sabe de x_1 causa un cambio en lo que se sabe de la variable aleatoria x_2. Este cambio es normalmente el resultado de nueva información que llega sobre x_1; a esta nueva información se le llama evidencia. Esta relación de causalidad entre las variables da nombre a los nodos; de esta forma, el nodo x_1 se llama «padre» y el nodo x_2, «hijo».La relación causal que existe entre las variables aleatorias x_1 y x_2 implica que la distribución conjunta puede ser expresada como un producto de probabilidades, es decir, $P(x_1)\,P(x_2|x_1)$, que no es más que la relación básica de probabilidad en forma gráfica. En general, los grafos dirigidos forman parte de una red de nodos que conectan variables mediante algún tipo de relación.
b. Cada variable tiene asociado un conjunto finito de estados mutuamente excluyentes.
c. Las variables junto con los arcos o segmentos dirigidos forman un grafo acíclico dirigido (GAD).Las RB contienen relaciones de causalidad y, por lo tanto, sus nodos están conectados por segmentos dirigidos. Son parte de un subconjunto de modelos de grafos conocidos como grafo acíclico dirigido (GAD). Los GAD son construidos con relaciones como las de la figura 2 como su bloque básico. Estos bloques están dispuestos de tal manera que son acíclicos, es decir, se mueven a lo largo de los bordes en las direcciones implicadas y es imposible volver a un nodo anterior. Asociado con cada GAD se tiene un conjunto de probabilidades condicionales

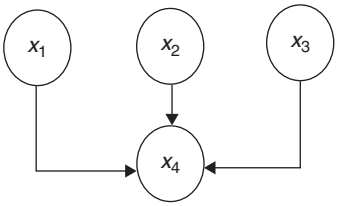

Figura 2. Grafo acíclico dirigido de 4 nodos.

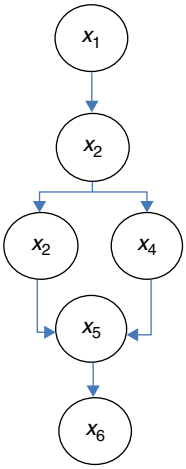

Figura 3. Ejemplo de una red bayesiana.

que señalan el comportamiento del nodo condicionado a sus «padres».El «padre» del nodo x_4 en la figura 2 es el conjunto de nodos $\{x_1, x_2, x_3\}$.

d. Para cada variable x_4 con «padres» $x_1, x_2 \ldots x_n$ existe una probabilidad asociada definida por $P(X_4|x_1, x_2 \ldots x_n)$. Si x_4 no tiene «padres», la probabilidad $P(x_4)$ es independiente.

Sea $X = \{x_1, x_2 \ldots x_n\}$ una variable aleatoria, si su función de distribución conjunta está definida por $P(X) = P(x_1, x_2 \ldots x_n)$. La función $P(X)$ crece exponencialmente en su complejidad con el número de variables. Las RB proporcionan una representación compacta de $P(X)$, factorizando la distribución conjunta en una distribución condicional local para cada variable dados sus «padres».

Sea $p_a(x_i)$ el conjunto de valores que toman los nodos «padres» de la variable x_i; entonces, la distribución conjunta total está dada por: $P(X) = P(x_1, x_2 \ldots x_n) = \prod_{i=1}^{n} P(x_i|p_a(x_i))$Por ejemplo, al analizar la figura 3, la probabilidad conjunta total (Cowell, Dawid, Verrall y Yoon (2007) está dada por:

$$P(X) = P(x_1, x_2, x_3, x_4, x_5, x_6) = \prod_{i=1}^{n=6} P(x_i|p_a(x_i))$$

Al mantener el supuesto de independencia en la construcción de la RB, el número de probabilidades condicionales a ser evaluadas se reduce considerablemente. Una RB se emplea básicamente

para inferencia, calculando las probabilidades condicionales, dada la información disponible hasta el momento para cada nodo.

La factorización de la distribución conjunta proviene de la propiedad de independencia condicional inherente a la estructura de los GAD. Sin embargo, esta propiedad puede hacer las manipulaciones de los GAD muy complejas, sobre todo si los nodos representan variables con muchos estados. Una factorización que simplifica los cálculos es la estructura llamada «árboles de derivación» *(Junction Tree)*. Esta estructura realiza cálculos modulares localizados que se ejecutan utilizando un algoritmo de «paso de mensajes». Estos árboles de derivación son grafos no direccionales y consisten en una colección de grafos también llamados universos de creencias; son grupos de nodos en el que cada nodo del grupo está conectado con cada otro nodo en el grupo (Cowell, Dawid, Luritzen y Spiegelhalter (1999).

El proceso de transformar un GAD en un árbol de derivación se llama triangulación, aun cuando el nombre hace referencia estrictamente a solo uno de los pasos. El proceso consta de 3 fases:

a. Primera fase: *moralización*. En este paso, todos los nodos padre de un nodo hijo en común que no están conectados son unidos con un segmento no dirigido; una vez hecho esto, a todos los segmentos se les quita la dirección.
b. Segunda fase: *triangulación*. Un ciclo es una secuencia de nodos conectados por segmentos que comienzan y terminan en el mismo nodo. Un ciclo de longitud *n* consiste de una secuencia de *n* segmentos consecutivos. La triangulación es el proceso de agregar segmentos no dirigidos de tal manera que cualquier ciclo que tiene una longitud no mayor a tres.
c. Tercera fase: *especificación del árbol de derivación*. Una vez obtenido el gráfico triangular, un árbol de derivación puede ser especificado.

Expresado el grafo de esta forma, la evidencia puede ser incorporada localmente a cada nodo con un número de cálculos menores. La información de los nodos actualizados es propagada por todos los nodos del árbol. Por lo tanto, no hay necesidad de utilizar las probabilidades conjuntas de todo el grafo; se hace localmente en cada nodo.

Una vez que se tiene el árbol de derivación es posible hacer inferencia a partir de él. Los pasos para la inferencia son:

a) Definir distribuciones *a priori*.

   Esta es la distribución no condicional *a priori* de los nodos, sin «padres», y la distribución condicional *a priori* para los nodos «hijos». Para cada función de distribución *a priori* se necesitaran las probabilidades de cada configuración de la combinación de estados de las variables involucradas.

   Para cada distribución *a priori* se necesitarán los datos para cada estado, lo cual puede ser determinado por:
   - Opinión de los expertos.
   - Estimación por máxima verosimilitud. El método de máxima verosimilitud asume que la información pasada es relevante y completa.

   En la práctica, no habrá una clara selección del método; los expertos también se basarán en los datos disponibles pero matizados con su experiencia y conocimiento respecto a su proyección para futuros eventos.
b) Inicializar el árbol. Incorporar la evidencia de cada nodo.

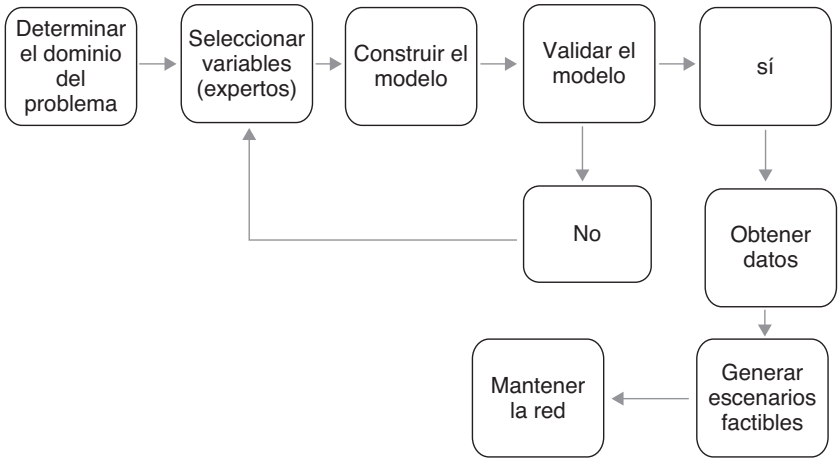

Figura 4. Diagrama de flujo de una red bayesiana.
Fuente: Elaboración propia.

c) Propagación. Enviar los mensajes hacia todos los nodos.

Para construir una RB se deben seguir los siguientes pasos (fig. 4):

- Definir el dominio del problema donde se especifique el propósito de la red.
- Identificar las variables o nodos importantes para el dominio del problema.
- Representar gráficamente la interrelación entre nodos o variables.
- Validar el modelo resultante con los expertos en el tema hasta lograr el consenso.
- Una vez validado y establecido el modelo, se cuantifica la red incorporando la opinión de los expertos, se crean escenarios factibles con la red y se mantiene actualizada a la red.

Construcción de la red bayesiana para el proceso de transacciones electrónicas

Construcción de la red

El proceso de construcción de una RB se divide en 2 grandes fases:

- La estructura del modelo.
- La cuantificación de la red.

Para llevar a cabo la estructura del modelo es necesario definir primero el dominio de la red, el cual consiste en diseñar un modelo que permita administrar el RO para el proceso de transacciones electrónicas.

El siguiente paso es identificar las variables o nodos importantes y, por último, definir las interrelaciones entre nodos para realizar su representación gráfica.

Tabla 1
Factores de riesgo

Grupo	Factores de riesgo
Datos	Control de acceso a la base de datos
	Respaldo de la base de datos (pérdida de la información)
	Robo o fraude
Aplicativos	Disponibilidad sitio web
	Fallas de sistemas
	Inactividad de las transacciones
Gestión	Errores humanos
	Aspectos normativos
Infraestructuras	Fallas o interrupción del suministro eléctrico
	Firewall
	Calidad del hardware
	Factores externos

Identificación de los nodos

Para la identificación de los nodos es necesario detallar los factores de riesgo asociados al proceso de transacciones electrónicas en la empresa. Los factores de riesgo son medidas que impactan el perfil de riesgo de las organizaciones; cambios en estas medidas implican cambios en el perfil de riesgo y son fundamentales para el proceso de gestión de riesgo en la organización.

Para llevar a cabo esta parte del proceso se realizaron varias sesiones con un grupo de expertos en las áreas de riesgo, operaciones, sistemas y costos en empresas financieras. Los factores de riesgo identificados para el proceso de transacciones electrónicas se agruparon en 4 categorías: datos, aplicativos, gestión e infraestructuras, y se presentan en la tabla 1:

Todos los factores de riesgo identificados son candidatos a emplearse como nodos en la construcción de la RB.

El siguiente paso es dar la descripción detallada para evitar ambigüedades en su uso y establecer sus estados o posibles valores (tabla 2).

Estructura de la red bayesiana

Para construir la RB se considera a los factores de riesgo seleccionados como nodos, los cuales serán conectados mediante arcos dirigidos para formar una estructura causal que muestre la dependencia entre estos. La estructura de la red se va configurando mediante un proceso iterativo. Se inicia con una red muy sencilla y se van agregando o eliminando nodos dependiendo de la información obtenida en las sesiones con los expertos, hasta que se obtiene una red aprobada por todos ellos.

Dada la complejidad del proceso, la RB se dividió en 2 subredes: una que modelara la frecuencia y otra la severidad; una vez obtenidos los resultados de la propagación de cada una de ellas de manera individual, estas se agregan a través del método de simulación Montecarlo para obtener el monto la pérdida esperada por RO para el proceso analizado.

Tabla 2
Descripción de los nodos

Nodo	Descripción
Control de acceso a la base de datos	Administración de los usuarios con permisos a la base de datos
Respaldo de la base de datos (pérdida de la información)	Asegurar y/o resguardar la información generada por los procesos operativos de tal manera que se mantenga íntegra y disponible
Robo o fraude	Información extraída ilegalmente de la base de datos por gente externa o interna en la organización
Disponibilidad sitio web	Disponibilidad que los servicios en línea
Fallas en los sistemas	Incidencias que ponen en riesgo la funcionalidad de la cadena productiva
Inactividad de las transacciones	Número de transacciones no procesadas
Errores humanos	Operación manual del personal con los conocimientos o aptitudes necesarias para el desempeño de sus funciones asignadas en la operación de los procesos o atención de servicios de transacciones electrónicas
Aspectos normativos	Disposiciones o resoluciones de las autoridades con efectos adversos o desfavorables en la operación como fuente de RO en la gestión de procesos o modificación o creación de aplicativos
Fallas o interrupción del suministro eléctrico	Discontinuidad en la operación debido a falta de energía
Firewall	Discontinuidad en la operación o afectación a la integridad de la base de datos debido a vulneración del sistema por falta de protección preestablecida
Calidad del hardware	Calidad de los componentes del hardware
Factores externos	Posibilidad o exposición a sufrir afectaciones en los activos de la empresa, causadas por fuentes externas o internas de manera intencional o no

Frecuencia

El nodo «inactividad de las transacciones» muestra el número de transacciones del proceso de transacciones electrónicas no procesadas y puede verse afectado por un deficiente respaldo de la base de datos, que a su vez es consecuencia de una deficiente calidad del hardware de la empresa. Otro motivo es la falta de disponibilidad del sitio web, ya que todas las transacciones son recibidas en línea y esto puede verse afectado por fallas en el suministro de energía eléctrica. La última causa de inactividad de las transacciones en el proceso analizado es el robo o fraude en la base de datos, que tiene como origen un deficiente control de acceso a la base de datos que permita que personal no calificado de la institución modifique o altere los registros o bien vulnerabilidad de la base de datos por falta o fallas en la protección de acceso a la información para personas ajenas a la institución.

La figura 5 muestra la red completa correspondiente al proceso de transacciones electrónicas desde el enfoque de la frecuencia. Se genera a través de los nodos identificados en el proceso. El modelo ha sido definido para reflejar incidentes en un periodo de un día.

Los posibles estados para cada nodo de la red de frecuencia se exponen en la tabla 3.

El nodo «fallas o interrupción del suministro eléctrico» mide el número de veces que se tuvieron fallas de suministro de electricidad; «Firewall» si se cuenta o no con un sistema de protección contra intrusos; todos los nodos anteriores son variables dicotómicas con estados «sí» o «no».

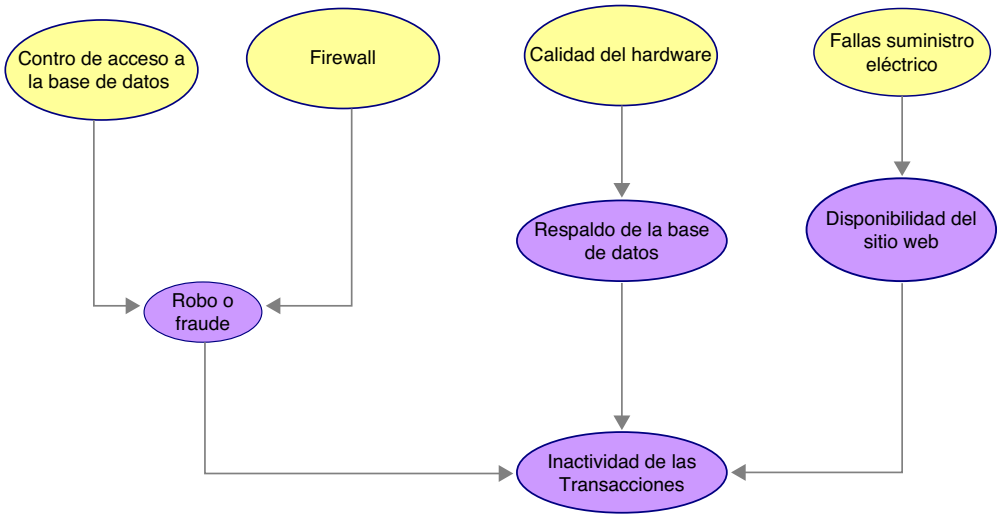

Figura 5. Red de frecuencia.

Tabla 3
Estados de los nodos de frecuencia

Nodos	Estado
Control de acceso a la base de datos	Alto
	Bajo
Respaldo de la base de datos (pérdida de la información)	100%
	90%
	80%
Robo o fraude	Sí
	No
Disponibilidad sitio web	Sí
	No
Inactividad de las operaciones en línea	0, 1, 2 . . . 10
Fallas o interrupción del suministro eléctrico	Sí
	No
Firewall	Sí
	No
Calidad del hardware	Alta
	Media

En nodo «control de acceso a la base de datos» califica si se tiene un alto o bajo control de acceso a la base de datos por parte de los usuarios dentro de la organización, clasificados en «alto» o «bajo».

En «respaldo de la base de datos (pérdida de la información)» se mide el porcentaje de respaldo actualizado que se tiene de la base de datos al cierre de la operación.

«Robo o fraude» registra el número de incidencias por estos eventos. «Disponibilidad sitio web» si se tiene o no acceso en línea a las transacciones electrónicas, e «inactividad de las transacciones» el número de transacciones no procesadas en un periodo de tiempo de un día.

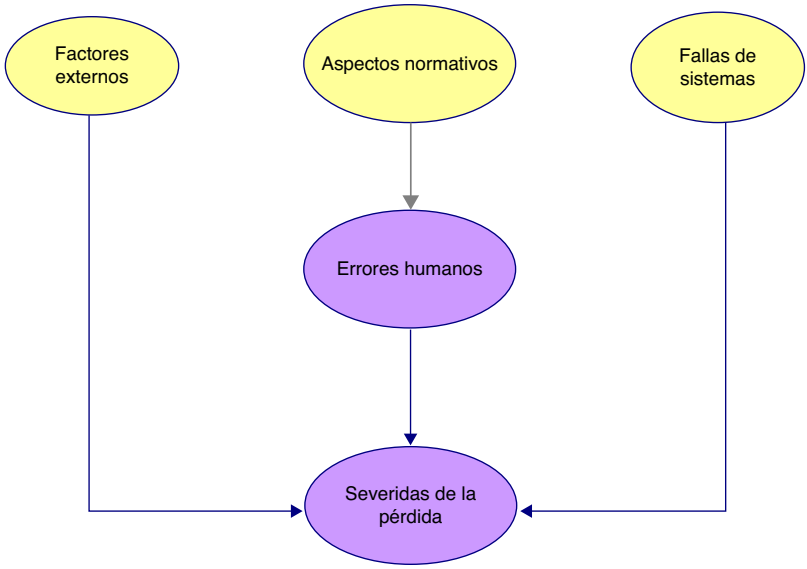

Figura 6. Red de severidad.

Severidad

El nodo «severidad de la pérdida» es la pérdida económica generada por fallas en los sistemas, afectaciones a los activos de la empresa causadas por agentes internos o externos a la misma o bien por errores humanos derivados principalmente de modificaciones o nuevas disposiciones emitidas por la entidad reguladora.

La figura 6 muestra la red completa correspondiente al proceso de transacciones electrónicas para la severidad. Se genera a través de los nodos identificados en el proceso. El modelo ha sido definido para reflejar pérdidas económicas en un periodo de un mes.

Los posibles estados para cada nodo de la red de frecuencia se indican en la tabla 4.

Todos los nodos de la red de severidad están medidos en términos de la pérdida económica que representa alguna falla en estas variables de riesgo.

Probabilidad condicional

Cada nodo representa una variable aleatoria discreta con un número finito de estados, o podría ser también a través de una variable continua. Si el nodo es discreto, se asocian probabilidades a los diferentes posibles estados. Si el nodo es continuo se asocia una función de densidad; en este trabajo se utilizaron solo variables discretas.

Cada nodo de la red tiene asociada una tabla de probabilidad condicional que determina el nivel de interrelación de los nodos. Estas fueron construidas a partir de información subjetiva obtenida de la opinión de los expertos en el proceso. Las tablas de probabilidad condicional para los nodos que conforman la red de frecuencia y severidad se presentan más adelante.

En la siguiente sección se procederá a describir las interacciones entre los nodos, así como las probabilidades asignadas *a priori* con base fundamentalmente en la opinión de expertos. También se plantea el desarrollo del modelo empleado y se realiza la propagación del modelo para obtener

Tabla 4
Estados de los nodos de severidad

Nodos	Estado
Fallas de sistemas	[$0, $20,000)
	[$20,000 - $50,000]
	Más de $50,000
Errores humanos	[$0 - $5,000)
	[$5,000 - $10,000]
	Más de $10,000
Aspectos normativos	[$0 - $3,000)
	[$3,000 - $15,000]
	Más de $15,000
Factores externos	[$0 -$10,000)
	[$10,000 - $50,000]
	Más de $50,000
Severidad de la pérdida	[$0 - $8,500)
	[$8,500 - $20,000]
	($20,000–$40,500]
	Más de $40,500

las probabilidades *a posteriori* y con ello establecer una medida de capital de riesgo para el proceso analizado.

Resultados obtenidos al ejecutar el modelo de redes bayesianas

Cuantificar la RB implica asignar una distribución de probabilidad para cada nodo de la red. La información para la asignación de probabilidades puede ser obtenida de datos estadísticos si están disponibles, o bien del juicio de los expertos. La confianza en la precisión de los consejos de los expertos es un requisito indispensable para utilizar este método. Ambas fuentes de información pueden ser empleadas separadas o conjuntamente.

En el caso que se aborda en este trabajo se utilizó solo la opinión de los expertos, datos subjetivos para todos los nodos. Una vez construida la RB, esta puede ser utilizada para realizar un análisis de sensibilidad de las predicciones o conclusiones respecto a los supuestos iniciales.

Cuantificación de la red bayesiana para el proceso de transacciones electrónicas

Este proceso implica obtener los datos para determinar la distribución de probabilidades de cada nodo de la red. En la segunda sección se explicó que existen 2 formas de definir las distribuciones *a priori* de los nodos: opinión de los expertos o estimación por máxima verosimilitud, para la cual se necesita información de los nodos completa. En este trabajo, para la cuantificación de la red se consultó principalmente la opinión de los expertos de diferentes áreas y se llegó a la definición de las probabilidades *a priori* para cada nodo.

A continuación se presentan los resultados obtenidos.

Obtención de datos

En general, en este tipo de riesgo los datos históricos disponibles son escasos y de difícil codificación para incorporarse a la RB, por lo que la cuantificación de la red se realizó con datos

subjetivos obtenidos de los expertos tanto para la frecuencia como para la severidad. Una de las ventajas del enfoque bayesiano es que permite la incorporación de la opinión de los expertos para obtener las probabilidades marginales o condicionales que alimentan el modelo.

Posteriormente, una vez que se obtenga la evidencia muestral, la distribución *a priori* es modificada con la nueva información y surge la distribución a posteriori. Con ella se formulan inferencias con respecto al parámetro.

Con las distribuciones *a priori* definidas para cada nodo en la sección anterior es posible inicializar el árbol y posteriormente realizar la propagación de la información para obtener las probabilidades *a posteriori*. La distribución a posteriori es la probabilidad de que el parámetro θ tome cierto valor, dadas las observaciones X.

Probabilidades a priori

El parámetro de la función de distribución *a priori* es visto como una variable aleatoria a la que, antes de la evidencia muestral, se le asigna una distribución *a priori* con base a un nivel de percepción del experto con respecto al comportamiento del parámetro aleatorio.

En caso de contar con información de datos históricos sobre las variables de interés, se ajusta la distribución de probabilidades correspondiente y se calculan las probabilidades requeridas.

La distribución de mayor uso para el ajuste de variables de frecuencia para este tipo de casos es la Poisson. Sin embargo, también suelen emplearse las distribuciones de tipo binomial, binomial Weibull y negativa (Bülmann y Gisler, 2005). Para las variables de severidad la distribución de mayor uso es la Lognormal, y pueden emplearse también las distribuciones Exponencial, Gamma, Beta, Pareto y Weibull (Bülmann y Gisler, 2005). En seguida se realizan estimaciones para encontrar el mejor parámetro; se realizan pruebas para evaluar la calidad del ajuste y finalmente obtener una curva a la distribución de pérdidas totales obtenida; para la comprobación de la bondad de ajuste se emplean técnicas estadísticas estándar, como el test de Pearson, Chi-cuadrado y la prueba de Kolmogorov-Smirnov.

Probabilidades a priori *para la frecuencia*

Para cada nodo, las probabilidades *a priori* definidas para la red de frecuencia con base en el juicio de los expertos se presentan a continuación.

Control de acceso a la base de datos (CABD). Es el nivel seguridad que se tiene por parte de la empresa para que los empleados usuarios de la base de datos accedan a ella. Se definieron 2 niveles: alto y bajo. En este tipo de empresas la confidencialidad de la base de datos es de vital trascendencia, por lo cual su resguardo y administración obliga a disponer de claves de acceso restringido para ingresar a ella.

Se definió con una probabilidad del 90% que el acceso a la misma es de seguridad alta, y con 10% de probabilidad se tiene un nivel de seguridad bajo.

Por lo tanto, la tabla de probabilidades *a priori* para este nodo es la siguiente:

Control de acceso a la base de datos	
Estado	Probabilidad
Alta	0.90
Baja	0.10

Firewall. Es un software o hardware que comprueba la información procedente de Internet o de una red y bloquea o permite el paso de esta al equipo. Un firewall puede ayudar a impedir que hackers o software malintencionado obtengan acceso al equipo a través de una red o de Internet.

Para este nodo se definen 2 estados posibles: contar con Firewall o no hacerlo. Es la función de la empresa administrar la base de datos, y por ello es responsable de su seguridad. Por tal motivo, la implementación del Firewall es prioridad para las instituciones; en general se cuenta con Firewall actualizado con una probabilidad del 95%. La tabla de probabilidades *a priori* para este nodo es la siguiente:

Firewall	
Estado	Probabilidad
Sí	0.95
No	0.05

Robo o fraude. Información extraída ilegalmente de la base de datos por gente externa o interna en la organización. Este nodo tiene como padres el control de acceso a la base de datos (CABD) y el Firewall. Considerando la opinión de los expertos, se definen 2 posibles estados para esta variable: que exista robo o fraude derivado de un deficiente control de acceso a la base de datos y/o inexistencia de Firewall. La tabla de probabilidad condicional *a priori* para este nodo de la red de frecuencia es la siguiente:

	Robo o fraude			
CABD	Alta		Baja	
Firewall	Sí	No	Sí	No
Sí	0.03	0.50	0.25	1.00
No	0.97	0.50	0.75	–

Dada una alta calidad en el control de acceso a la base de datos y un Firewall instalado, la probabilidad de sufrir un robo o fraude de la base de datos es del 3%. En el caso de no tener controles de acceso a la base de datos y carecer de Firewall, la probabilidad de robo o fraude es del 100%.

Calidad de hardware (CH). Calidad en los componentes del hardware empleados en la administración de la base de datos. Los expertos del área de sistemas consideran que existe una continua actualización de los componentes de hardware, y con base en ello se definieron 3 posibles estados: contar con hardware de la más alta calidad con una probabilidad del 80%; que sea de mediana calidad pero funcional para las necesidades de la empresa con probabilidad del 15%, o bien que se vuelva equipo obsoleto que retrase o imposibilite la operación con probabilidad del 5%.

Calidad del hardware	
Estado	Probabilidad
Alto	0.80
Medio	0.15
Bajo	0.05

Respaldo de la base de datos (RBD). Como parte de las medidas de prevención es imprescindible realizar respaldos a la base de datos con el objetivo de asegurar y/o resguardar la información generada por los procesos operativos de tal manera que se mantenga íntegra y disponible. Para este ejemplo se consideró que dicho respaldo se realiza semanalmente. Derivado de la calidad del hardware, el respaldo se puede ver afectado, por lo que los expertos de operaciones sugieren 3 posibles niveles: respaldo de la información realizado con éxito al 100%; respaldo del 90%, o

solo del 80%. La tabla de probabilidad condicional *a priori* para este nodo de la red de frecuencia es la siguiente:

	Respaldo de la base de datos (RBD)		
CH	Alto	Medio	Bajo
100%	0.95	0.75	0.45
90%	0.05	0.15	0.30
80%	0	0.1	0.25

La tabla anterior se interpreta de la siguiente manera: la probabilidad de contar con un respaldo de la base de datos del 100% dada una alta calidad del hardware es del 95%, y con probabilidad del 5% el respaldo de la base de datos será del 90%. Si la calidad del hardware es «media», con probabilidad del 75% el respaldo de la base de datos se hará al 100%; con probabilidad del 15% el respaldo solo será del 90%. Por último, si la calidad del hardware es baja, con probabilidad del 45% la base de datos se respalda solo al 100%, con probabilidad del 30% será solo respaldada al 90%, y con probabilidad del 25% el respaldo de la información será solo del 80%.

Fallas de suministro eléctrico (FSE). Discontinuidad en la operación debido a falta de energía eléctrica. Los estados definidos son obtenidos de la experiencia de los últimos años por parte de los expertos, quienes consideraron que las fallas en el suministro eléctrico causaron en el 20% de los eventos interrupción de la operación. La tabla de probabilidad *a priori* es la siguiente:

Fallas del suministro eléctrico (FSE)	
Estado	Probabilidad
Sí	0.20
No	0.80

Disponibilidad del sitio web (DSWEB). La disponibilidad de los servicios en línea puede verse afectada directamente por la interrupción del suministro de energía eléctrica. En caso de que no exista energía eléctrica, los servicios en línea se ven interrumpidos con probabilidad del 95%. En caso de que no haya fallas en el suministro de electricidad con una probabilidad del 95% el sitio web está disponible para la comunicación en línea. La tabla de probabilidad condicional asociada a este nodo es la siguiente:

	Disponibilidad del sitio web (DSWEB)	
FSE	Sí	No
Sí	0.05	0.95
No	0.95	0.05

Inactividad de las operaciones en línea (IOL). Es el nodo objetivo y se define como el número de transacciones no concluidas. Sus estados son de cero a diez. La inactividad de las transacciones es consecuencia de la disponibilidad del sitio web, el porcentaje de respaldo de la base de datos y de haber sufrido robo o fraude de la información; la transacción fue enviada pero no recibida para procesar la solicitud. La tabla 5 muestra las probabilidades condicionales asociadas a este nodo.

Si el sitio web se encuentra disponible, el respaldo de la base de datos se realizó al 100% y no hubo robo o fraude, con probabilidad del 45% el número de operaciones en línea no procesadas es cero, con 22% de probabilidad será una la transacción no procesada, con 14% de probabilidad

Tabla 5
Probabilidades de la inactividad de las operaciones en línea

Inactividad de las operaciones en línea (IOL)												
DSWEB	Sí						No					
RBD	100%		90%		80%		100%		90%		80%	
Robo o Fraude	Sí	No	Sí	No	Sí	No	Sí	No	Sí	No	Sí	No
0	0.13	0.45	0.12	0.38	0.05	0.19	0.10	0.25	0.04	0.17	0.00	0.07
1	0.30	0.22	0.29	0.27	0.30	0.35	0.21	0.30	0.22	0.32	0.22	0.34
2	0.19	0.14	0.23	0.14	0.24	0.16	0.28	0.22	0.30	0.24	0.34	0.25
3	0.19	0.08	0.22	0.08	0.23	0.09	0.25	0.06	0.26	0.06	0.16	0.07
4	0.06	0.05	0.04	0.05	0.05	0.06	0.05	0.06	0.05	0.06	0.09	0.06
5	0.06	0.04	0.03	0.04	0.04	0.06	0.03	0.04	0.03	0.04	0.06	0.06
6	0.03	0.01	0.03	0.01	0.03	0.03	0.03	0.02	0.03	0.04	0.04	0.04
7	0.02	0.01	0.02	0.01	0.02	0.02	0.03	0.02	0.03	0.03	0.03	0.04
8	0.02	0.00	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.03
9	0.00	0.00	0.00	0.00	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.03
10	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.01

2 operaciones en línea, con 8% de probabilidad 3 operaciones en línea; 4 operaciones en línea no procesadas con 5%, y 5 o más con probabilidad de 6%.

En caso de que el sitio web se encuentre disponible, el respaldo de la base de datos se haya efectuado al 100% pero haya robo o fraude, la probabilidad de que cero operaciones en línea no sean procesadas es del 13%, con probabilidad del 30% una transacción no será procesada; con probabilidad del 19% 2 operaciones en línea no serán procesadas, y el mismo porcentaje para 3 operaciones en línea. Que 4 operaciones en línea no sean procesadas sucede con probabilidad del 6%, lo mismo para 5 operaciones en línea, y con probabilidad del 7% 6 o más operaciones en línea no logran procesarse.

Las otras probabilidades condicionales se leen de manera similar en la tabla 5.

Probabilidades a priori *para la severidad*

Para cada nodo, las probabilidades *a priori* definidas para la red de severidad con base en el juicio de los expertos se presentan a continuación.

Factores externos. Se entiende como la posibilidad o exposición a sufrir afectaciones en los activos de la empresa, causadas por fuentes externas o internas de manera intencional o no como es el caso de manifestaciones u atentados, entre otros. Estos eventos podrían ocasionar pérdidas económicas para la organización clasificadas en montos menores a $10,000; entre $10,000 y $50,000, o más de $50,000.

La tabla de probabilidades asociada a este nodo es la siguiente:

Factores externos	
Estado	Probabilidad
< 10,000	0.65
10,000-50,000	0.28
> 50,000	0.07

La probabilidad de que ocurra una pérdida por un importe menor a $10,000 derivado de un factor externo a la organización es del 65%, con probabilidad del 28% implica una erogación entre $10,000 y $50,000, y con probabilidad del 7% la pérdida es mayor a $50,000.

Aspectos normativos. Disposiciones o resoluciones de las autoridades con efectos adversos o desfavorables en la operación como fuente de RO en la gestión de procesos, modificación de los mismos o creación de aplicativos[2]. No acatar en el tiempo debido las disposiciones representa multas o sanciones establecidas por la autoridad para las instituciones financieras. Los montos de multas y sanciones se pueden agrupar en 3 intervalos: pagos menores a $3,000, entre $3,000 y $15,000, y mayores a $15,000.

Las probabilidades asignadas para este nodo se muestran en la siguiente tabla de probabilidad.

Aspectos normativos	
Estado	Probabilidad
< 3,000	0.78
3,000-15,000	0.17
> 15,000	0.05

Existe la probabilidad del 78% de que un aspecto normativo implique el pago a la autoridad por un monto menor a $3,000, con probabilidad del 17% la sanción sería entre $3,000 y $15,000, y con probabilidad del 5% el monto es mayor a $15,000.

Errores humanos. Operación manual del personal con los conocimientos o aptitudes necesarias para el desempeño de sus funciones asignadas en la operación de los procesos. Fallas derivadas de la falta de conocimiento del personal en materia de las disposiciones emitidas por la entidad reguladora se traducen en pérdidas económicas para la institución. De acuerdo a la opinión de los expertos es difícil registrar a detalle estos datos. Al igual que para los otros nodos, se recurrió al juicio de los expertos para la definición de las probabilidades condicionales *a priori* asociadas a este nodo.

	Errores humanos		
Aspectos normativos	< 3,000	3,000-15,000	> 15,000
< 5,000	0.70	0.43	0.20
5,000-10,000	0.25	0.39	0.58
> 10,000	0.05	0.18	0.22

Derivado de un aspecto normativo cuya sanción fue menor a $3,000, con probabilidad del 70% se comete un error humano que implique una pérdida por un monto menor a $5,000; con una probabilidad del 25% la pérdida se ubica entre $5,000 y $10,000, y con una probabilidad del 5% la pérdida es mayor a $10,000. Si el monto de pérdida derivado de un aspecto normativo se encuentra entre $3,000 y $15,000, con probabilidad del 43% el monto de la pérdida por un error humano es menor a $5,000, con probabilidad del 39% el monto está entre $5,000 y $10,000, y con probabilidad del 18% es de más de $10,000. Si el monto de la pérdida derivada de aspectos normativos es superior a los $15,000, con probabilidad de 20% el error humano implicará una pérdida menor a $5,000, con probabilidad de 58% la pérdida estará entre $5,000 y $10,000, y con probabilidad de 22% el monto de pérdida consecuencia de un error humano será mayor a $10,000.

[2] Un aplicativo es el desarrollo de software de una aplicación.

Fallas de sistemas. Las incidencias que ponen en riesgo la funcionalidad de la cadena productiva y la solución a los mismos representan desembolsos económicos para la empresa. A pesar de tener registro de los incidentes ocurridos no se conoce el monto de la pérdida por cada uno de ellos y no se tiene identificado en la base de datos a cuál de los 4 procesos de la empresa afectó el incidente; por lo tanto, fue necesario recurrir a la opinión de los expertos en sistemas para determinar los montos potenciales y clasificar las pérdidas en 3 rangos: menores de $20,000; entre $20,000 y $50,000, y más de $50,000.

Fallas en sistemas	
Estado	Probabilidad
< 20,000	0.83
20,001-50,000	0.09
> 50,000	0.08

Con probabilidad del 83% la pérdida económica derivada de una falla en los sistemas es de monto menor a $20,000; con probabilidad del 9% la pérdida corresponde a un monto entre $20,000 y $50,000, y con probabilidad del 8% la pérdida es mayor a $50,000.

Severidad de la pérdida. Es el nodo objetivo de la red de severidad y representa la suma de las pérdidas asociadas con los nodos «factores externos», «fallas en sistemas» y «errores humanos», que a su vez tiene como causa los «aspectos normativos». La tabla de probabilidades condicionales asociadas a este nodo final es la siguiente.

Severidad de la pérdida									
Factores externos	< 10,000								
Errores humanos	< 5,000			5,000-10,000			> 10,000		
Fallas en sistemas	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000
0-8,500	1.00	0.60	0.10	0.95	0.50	–	0.90	0.35	0.02
8,500-20,000	–	0.30	0.30	0.03	0.30	0.20	0.05	0.25	0.10
20,000-40,500	–	0.10	0.20	0.02	0.20	0.30	0.03	0.15	0.20
Más de 40,500	–	–	0.40	–	–	0.50	0.02	0.25	0.68
Factores externos	10,000-50,000								
Errores humanos	< 5,000			5,000-10,000			> 10,000		
Fallas en sistemas	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000
0-8,500	0.93	0.50	0.01	0.80	0.30	–	0.70	0.25	–
8,500-20,000	0.05	0.30	0.15	0.12	0.10	0.10	0.15	0.15	0.10
20,000-40,500	0.02	0.20	0.40	0.05	0.45	0.30	0.10	0.40	0.10
Más de 40,500	–	–	0.44	0.03	0.15	0.60	0.05	0.20	0.85
Factores externos	> 50,000								
Errores humanos	< 5,000			5,000-10,000			> 10,000		
Fallas en sistemas	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000	< 20,000	20,000-50,000	> 50,000
0-8,500	0.60	0.20	–	0.57	0.25	–	0.45	–	–
8,500-20,000	0.13	0.10	0.12	0.12	0.15	0.05	0.17	–	–
20,000-40,500	0.18	0.40	0.22	0.20	0.50	0.20	0.25	0.50	0.05
Más de 40,500	0.09	0.30	0.66	0.11	0.10	0.75	0.13	0.50	0.95

Para un monto de pérdida menor a $10,000 como consecuencia de factores externos, una pérdida por error humano menor a $5,000 y pérdidas menores a $20,000 por fallas en los sistemas, con probabilidad del 100% se estima una pérdida económica menor a $8,500.

Dado un monto de pérdida entre $10,000 y $50,000 como consecuencia de un factor externo, una pérdida por error humano mayor a $10,000 y fallas en los sistemas por menos de $20,000, con probabilidad del 90% la pérdida será menor a $8,500; con probabilidad del 5% estará entre $8,500

y $20,000; con probabilidad del 3% será mayor a $20,000 y hasta $40,500, y con probabilidad del 2% será superior a 40,500

Las otras probabilidades condicionales de la tabla se leen de manera similar.

Probabilidades a posteriori

La utilización de la independencia condicional en la teoría de grafos y los recientes avances en el desarrollo de algoritmos para la propagación de la evidencia en estructuras gráficas ha hecho que los cálculos sean más sencillos al utilizar software disponible en versión libre para el diseño y propagación de las redes.

Para el cálculo de las probabilidades a posteriori de la RB se utilizó el algoritmo *Junction Tree* (árboles de derivación) por medio de un software de acceso libre llamado GeNIe versión 2.0[3]. El *Junction Tree* permite convertir el grafo dirigido acíclico en un árbol cuyos nodos son cerrados para después propagar, como se describió en la segunda sección.

A continuación se presentan las distribuciones a posteriori para cada uno de los nodos, tanto de la red de frecuencia como de la red de severidad.

Probabilidades a posteriori para la frecuencia

La figura 7 muestra un diagrama elaborado en GeNIe para las probabilidades *a posteriori* para los nodos de la red de frecuencia después de la propagación.

Posterior a la propagación, las probabilidades a posteriori para los nodos «padre» en la red no tienen cambio con respecto a las probabilidades *a priori*, debido a que en ellos no existe un proceso de propagación de mensajes.

Los resultados del nodo «robo o fraude» indican que existe una probabilidad del 92% de que la institución no se vea amenazada por este evento y con probabilidad del 8% se verán afectadas por el robo o fraude de la información de la base de datos. Lo anterior condicionado a las condiciones descritas para el control de acceso a la base de datos y al Firewall de la empresa.

Con una probabilidad del 90% el respaldo de la base de datos se efectúa en su totalidad, con una probabilidad del 8% solo se lleva a cabo al 90%, y con probabilidad del 3% el respaldo de la información queda únicamente al 80%, condicionado lo anterior a la calidad del hardware.

La disponibilidad del sitio web tiene una probabilidad del 77% de estar disponible y de 23% de estar fuera de línea, condicionado al suministro de energía eléctrica.

Finalmente, la distribución de probabilidades del nodo «inactividad de las operaciones en línea» muestra una probabilidad del 37% de tener cero operaciones no completadas; una probabilidad del 25% de que una transacción no sea completada; con probabilidad del 16% 2 transacciones; con probabilidad del 9%, 3, y 12% de probabilidad de que las transacciones no completadas excedan de 3.

Probabilidades a posteriori *para la severidad*

La figura 8 muestra el diagrama de las probabilidades *a posteriori* para los nodos de la red de severidad posterior a la propagación.

[3] Dispo nible en: ht tp://genie.sis.pitt.edu/downloads.html.

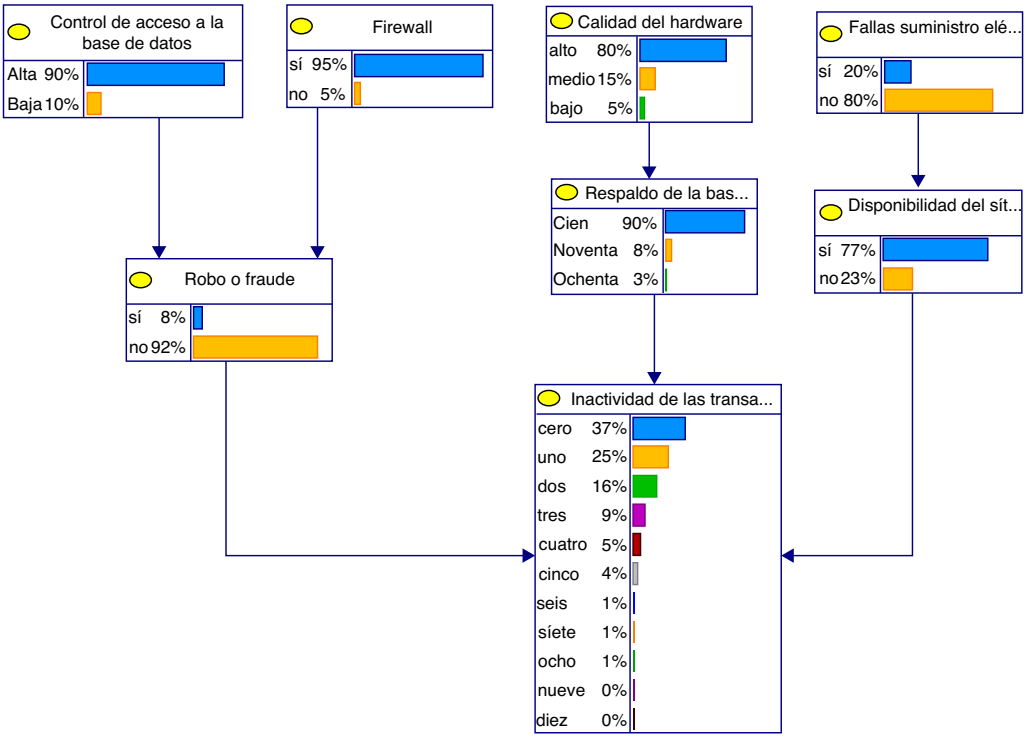

Figura 7. Diagrama en GeNIe de las probabilidades *a posteriori* para los nodos de la red de frecuencia.

Las pérdidas originadas por «errores humanos» como consecuencia de las disposiciones o modificaciones emitidas por la autoridad reguladora («aspectos normativos») tienen una probabilidad del 63% de ser menores a $5,000; con probabilidad del 29% la pérdida por este concepto estará entre $5,000 y $10,000, y con probabilidad del 8% será superior a $10,000. Lo anterior es consecuencia del alto nivel de atención que en general se tiene para que el personal operativo de empresas financieras se encuentre actualizado en las disposiciones emitidas por la entidad reguladora.

La distribución de probabilidades del nodo «severidad de la pérdida» tiene una probabilidad del 81% de que la pérdida sea menor a $8,500, una probabilidad del 7% de que esté entre $8,500 y $20,000; con probabilidad del 11% la pérdida es superior a $20,000, en un periodo de tiempo de 24 horas.

Cálculo del capital de riesgo operacional

Para el cálculo de capital de RO, la legislación a la que está sujeta la empresa, emitida por la entidad reguladora, no especifica un porcentaje sobre el cual debe estimarse la distribución de pérdida. La empresa ha definido que este sea del 95%.

Asumiendo que la frecuencia se distribuye $P(.|\lambda)$ y la severidad $f(.|\alpha)$, dados λ y α, y que la función de distribución *a posteriori* $\hat{\pi}(\lambda|N)$ y $\hat{\pi}(\alpha|X)$ para λ y α respectivamente son estimados con las funciones de distribución definidas con la opinión de los expertos y ponderadas con datos

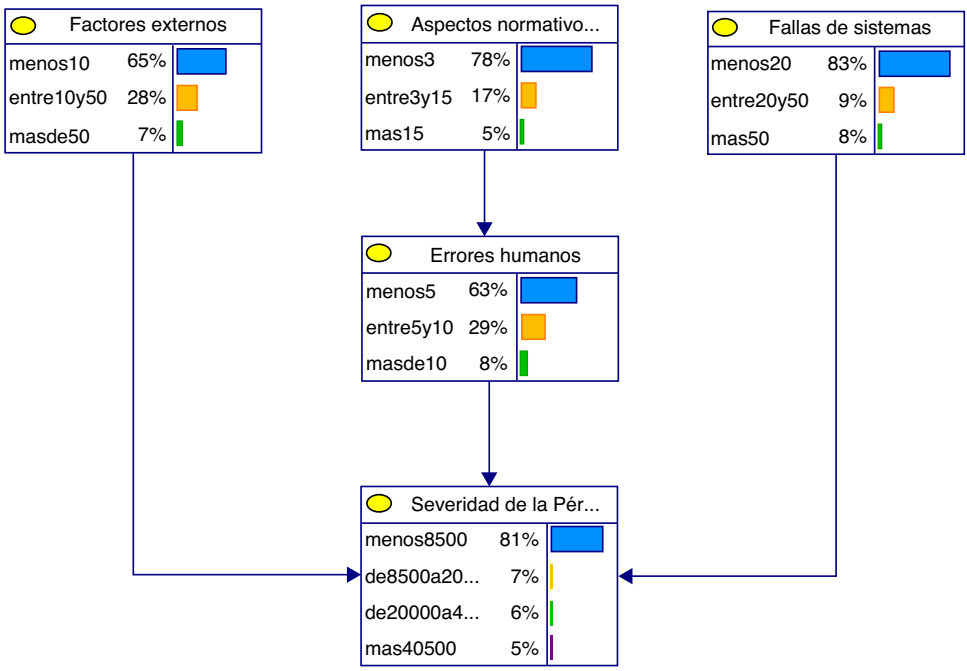

Figura 8. Diagrama en GeNIe de las probabilidades a posteriori para los nodos de la red de severidad.

históricos (cuando se tienen), la distribución de las pérdidas puede calcularse usando el método Montecarlo con los siguientes pasos:

1. Simular los parámetros λ y α para las distribuciones *a posteriori* $\hat{\pi}(\lambda|N)$ y $\hat{\pi}(\alpha|X)$.
2. Dado λ, simular un número de eventos N para distribución de la frecuencia.
3. Dado α, simular las severidades X_n para $n = 1, 2, \ldots, n$ para la distribución de la severidad $f(.|\alpha)$. Debe tenerse en cuenta que las severidades para el modelo se suponen independientes e idénticamente distribuidas para los parámetros dados λ y α.
4. Encontrar la pérdida esperada $Z = \sum_{i=1}^{N} X_n$.
5. Repetir los pasos 1 a 4 K veces para construir una muestra de las pérdidas $Z(K)$ con $k = 1, 2, \ldots, N$. Se ordenan los valores obtenidos en orden descendente y se calculan en el cuantil 0.95.

Valor en riesgo operacional para el proceso de transacciones electrónicas

Dada la carencia total de datos para el procedimiento en el cálculo del capital de RO del caso de análisis y contando solo con la opinión de los expertos, la RB tendrá que ser actualizada tan pronto se tenga un histórico de información disponible de los eventos en las variables de riesgo.

Bajo el supuesto de que el número de inactividad de las operaciones en línea esperada (N) es de 1.52 y la severidad de la pérdida esperada (X) de \$4,015, considerando independencia de las variables:

$$E(Pérdida) = E(N) * E(X) = 6,090$$

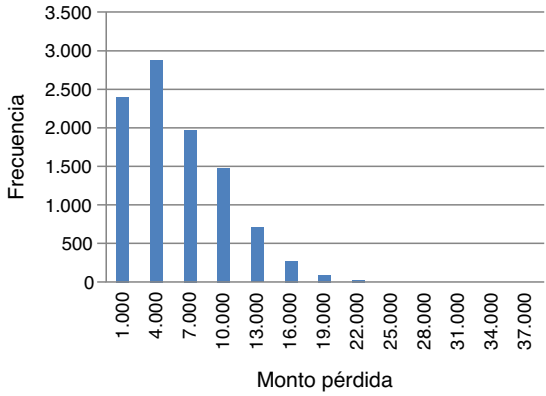

Figura 9. Pérdidas calculadas mediante simulación Montecarlo.

Para calcular la distribución de las pérdidas usando el método Montecarlo y siendo consistentes con la teoría presentada en la cuarta sección, se supone una distribución Poisson para la frecuencia de la inactividad de las transacciones con $\lambda = 1.52$ y una distribución Lognormal con $\mu = 4, 015$.

1. Para las distribuciones *a posteriori*, dada la falta de datos, los parámetros son los obtenidos por la red.
2. Se simularon 10,000 números posibles de eventos (N) para la distribución de frecuencia $P(N|\lambda = 1.52)$ utilizando el generador de números aleatorios Poisson de Excel.
3. Se simularon 10,000 números aleatorios para la distribución de la severidad Lognormal[4] $LN(X|\mu = 4, 015,\ \sigma = 652)$.
4. Obtener la pérdida esperada $E(P\acute{e}\ rdida) = E(N) * E(X)$.

Los valores obtenidos para las 10,000 simulaciones se ordenan en orden descendente y se obtiene el valor de la celda que corresponde al 95%; en este caso corresponde a un monto de OpVaR de $15,478.88.

La figura 9 muestra los resultados de pérdidas obtenidos del proceso de simulación Montecarlo.

El OpVaR con un nivel de confianza del 95% da una pérdida máxima esperada de $15,479 pesos diarios. De la cual la pérdida esperada es de $4,015 y la pérdida no esperada de $11,464 pesos diarios.

Es probable que el OpVaR calculado con el modelo bayesiano sea superior al calculado con el modelo clásico, lo cual se explica por la causalidad entre los distintos factores de riesgo que no considera el modelo clásico. Lo anterior representa una subvaluación de la máxima pérdida esperada bajo un enfoque clásico y hace que el modelo bayesiano sea una alternativa válida para la estimación del capital de RO.

Validación del modelo

El siguiente paso será la validación del modelo, lo cual deberá realizarse por expertos del área de riesgos y de operaciones de la empresa financiera en la que se aplique el modelo de RB para

[4] DIstr.Log.Inv(Aleatorio(),μ,σ)

el cálculo del RO. Una de las principales ventajas de las RB es que permite la actualización de datos conforme estos se obtienen para calibrar el modelo a la luz de la información disponible. Lo anterior permitirá una estimación precisa del capital de RO de acuerdo a las variables definidas y sus relaciones de causalidad, dada la información particular de cada institución.

Conclusiones

La limitante de encontrar bases de datos completas que contemplen los diferentes aspectos del RO y una compleja interacción entre las variables de riesgo identificadas en los distintos procesos de la empresa que lo caracterizan, hace que la medición de este riesgo requiera de técnicas dinámicas de medición.

Un enfoque bayesiano es una alternativa viable para el análisis de riesgos en condiciones de información insuficiente. Los modelos bayesianos son modelos causa-efecto dinámicos que incorporan información inicial a través de una distribución de probabilidad *a priori*, mediante la cual se puede incluir información subjetiva en la toma de decisiones, como la opinión de expertos, el juicio de analistas o las creencias de especialistas posicionados en cada uno de los procesos que contribuyen a la obtención de un producto o servicio que ofrece la empresa. Asimismo, la dinámica de la información permitirá la actualización del modelo incorporando la información más reciente y, en consecuencia, un análisis de riesgos actualizado para la toma de decisiones.

Un diseño exitoso de una RB se fundamenta en la descomposición significativa de un dominio del problema en un conjunto de proposiciones causales o condicionales sobre el dominio, lo que implica un pleno conocimiento de cada uno de los procesos que intervienen en el origen de un producto o servicio.

El uso de RB permite realizar un análisis de escenarios; es decir, si la RB se ejecuta, el efecto de los datos ingresados dentro de uno o más nodos son propagados por toda la red, en todas las direcciones, y la distribución marginal de los nodos es actualizada y, en consecuencia, brinda información oportuna al usuario.

La metodología utilizada en RO a través de RB permite describir un conjunto de recomendaciones que mitigan el riesgo en los procesos de la empresa y que impactan en su productividad; así mismo proporciona un valor al OpVaR más preciso, pues considera elementos que los métodos tradicionales no contemplan.

El registro de datos estadísticos de los distintos factores de riesgo identificados en cada uno de los procesos de la empresa permitirá incorporar funciones de distribución explícitas en cada proceso para determinar las probabilidades requeridas y realizar una análisis más detallado del uso del modelo de RB apegado a su entorno operacional, permitiendo mitigar el riesgo inherente en los distintos procesos, en la que al día de hoy se utilizó solo la opinión de un grupo de expertos —datos subjetivos— para todos los nodos.

Referencias

Alexander, C. (2002). *Operational Risk Measurement: Advanced Approaches*. Reading, UK: ISMA Centre, University of Reading.

Aczel, A. y Sounderpandian, J. (2009). *Business Statistics* (séptima edición, pp. 692–788). Mac Graw Hill.

Cardozo Ojeda, E. y Arguello Fuentes, H. (2011). Aprendizaje estructural de redes bayesianas: un enfoque basado en puntaje y búsqueda. *Ciencia e Ingeniería Neogranadina*, *21*, 29–50.

Cowell, R. G., Dawid, A. P., Luritzen, S. L. y Spiegelhalter, D. J. (1999). Probabilistic networks and expert systems. *Sping 2000*, 1–13.

Cowell, R. G., Dawid, A. P., Verrall, R. J. y Yoon, Y. K. (2007). Modeling operational risk with Bayesian networks. *The Journal of Risk and Insurance*, *74*(4), 795–827.

Hossack, I. B., Pollard, J. H. y Zehnwirth, B. (1999). *Introductory Statistics with Applications in General Insurance*. Cambridge University Press.

Leippold, M. (2003). The Quantification of Operational Risk. *Social Science Research Network*.

Neil, M., Marquez, D. y Fenton, N. (2004). Bayesian Networks to model expected and unexpected operational losses. *Risk Analysis*, *25*(4).

Reimer, K. y Neu, P. (2003). *Functional Correlation Approach to Operational Risk in Banking Organizations*. Kings College London: Dresdner Bank AG.