

COMUNICACIONES DEL CIMAT

PE91-05 (Julio de 1991)

UN MODELO BASADO
EN CADENAS DE MARKOV
PARA DAÑO ACUMULADO

108

Miguel Nakamura

y

Francisco J. Pérez A.

108

CENTRO DE
INVESTIGACION EN
MATEMATICAS

Apartado Postal 402

36000 - Guanajuato, Gto.

México

Tel. (473)271-55

Fax (473)257-49

E-mail CIMAT@UNAMVM1.BITNET

UN MODELO BASADO EN CADENAS DE MARKOV PARA DAÑO ACUMULADO

Miguel Nakamura S. y Francisco J. Pérez A.

Centro de Investigación en Matemáticas
Apartado Postal 402, Guanajuato, Gto. 36000

Indice

1.	Introducción	2
2.	Conceptos Básicos de Probabilidad	3
3.	El Modelo Basado en una Cadena de Markov	8
4.	Uso del Programa de Cómputo	12
5.	Bibliografía	16

1. INTRODUCCION

En el presente reporte tenemos como objetivo el describir un modelo estadístico para representar los tiempos de vida de componentes mecánicos que sufren desgaste o deterioro debido a uso cíclico, y presentamos una posible manera de estimar el modelo cuando se cuenta con algunas observaciones prácticas acerca de la duración de algunos componentes. Describimos también algunas de las conclusiones que pueden extraerse con este tipo de análisis, así como el empleo de un programa de cómputo que ha sido desarrollado para este fin.

Específicamente, después de instalar una pieza o refacción, una cantidad de interés es el tiempo T que transcurre antes de que falle o tenga que ser retirada de servicio y deba ser reemplazada por una nueva. Son pocos los casos en los cuales el valor de T resulta conocido de antemano, es decir, en los cuales T es una cantidad determinística; en la mayoría de los casos, el valor de T es en efecto una cantidad aleatoria. Nos interesa por tanto, obtener una descripción probabilística sobre T , lo cual se establece en forma elemental en la Sección 2. Para un tratamiento más extenso sobre el tema de probabilidad y variables aleatorias, puede consultarse por ejemplo la referencia [4], mientras que para una introducción a la estadística, [1] y [3].

En este trabajo, consideramos solamente el caso especial en el que T se mide en unidades enteras o ciclos de uso, los cuales no tienen necesariamente que estar relacionadas con el concepto de tiempo, como sucede en los siguientes ejemplos:

A) T es el número de aterrizajes que registra una unidad de frenos en una aeronave antes de tener que ser reemplazada por haber sufrido desgaste.

B) T es el número de corridas que hace un autobús entre dos ciudades dadas antes de que tenga que reemplazarse alguna refacción.

C) T es el número de clientes que utilizan una máquina registradora antes de que tenga que reemplazarse el carrete de papel.

D) T es el número de veces que se abre y se cierra una cerradura antes de que ésta falle.

Otra limitación del tipo de modelos que consideramos en este trabajo es la suposición de que las refacciones siempre se instalan nuevas. Se tornan mucho más complejos los modelos en los cuales se permite que algunas refac-

ciones que reemplazan a otras dañadas tengan a su vez algunos ciclos de uso acumulados de antemano. El término **daño acumulado** se refiere a la situación, típica en problemas de ingeniería, en la cual la pieza "acumula" irreversiblemente grados de deterioro o desgaste hasta fallar (difícilmente una pieza en uso se "renueva" por sí sola). Para una exposición mucha más completa y detallada sobre los modelos de daño acumulado, incluyendo aquellos que no tienen las limitaciones aquí descritas, remitimos al lector a [2].

2. CONCEPTOS BASICOS DE PROBABILIDAD

Sin definir formalmente la probabilidad ni sus propiedades, esta sección se limita más bien a una interpretación intuitiva del concepto. Sea x un número real, y T un tiempo aleatorio de falla como el descrito en la Sección 1. Para la *probabilidad de que T no supere a x* , se utiliza la notación $P(T \leq x)$. Por ejemplo, si $x = 1000$, $P(T \leq 1000)$ denota la probabilidad de que el componente falle antes de cumplir 1001 ciclos de uso. La probabilidad cumple ser un número entre 0 y 1, y posee la interpretación siguiente: Si fuese posible observar el desempeño del mismo componente un número grande de veces N , $P(T \leq x)$ representa la proporción aproximada de veces que el componente fallaría antes de cumplir x ciclos. Es decir, el número de veces que el componente fallaría antes de x ciclos sería aproximadamente $N \cdot P(T \leq x)$, o bien, $P(T \leq x)$ es aproximadamente la *frecuencia relativa* con la que los componentes fallan antes de x ciclos. Por supuesto, esto solamente constituye una interpretación de la probabilidad; en la realidad no sería posible observar N veces fallar al mismo componente. La afirmación $P(T \leq x) = 0$ para algún valor de x , refleja el caso en que T *nunca* resulta ser menor o igual que x .

Claramente, si T es en efecto un tiempo de falla y $x < 0$, entonces $P(T \leq x) = 0$, debido a que siempre se obtiene $T > 0$. La función de x definida por

$$F_T(x) = P(T \leq x)$$

recibe el nombre de **Función de Distribución de la variable T** . En general, si T es una variable no-negativa, como lo es un tiempo de falla, entonces $F_T(x) = 0$ si $x < 0$.

La función de distribución posee varias propiedades:

- i) $\lim_{x \rightarrow -\infty} F_T(x) = 0$ cuando $x \rightarrow -\infty$, y $\lim_{x \rightarrow \infty} F_T(x) = 1$ cuando $x \rightarrow \infty$.

- ii) F_T es no decreciente, es decir, si $x < y$, entonces $F_T(x) \leq F_T(y)$.
- iii) F_T es continua por la derecha, es decir, $\lim_{h \rightarrow 0^+} F_T(x+h) = F_T(x)$ cuando $h > 0$ y $h \rightarrow 0$.

En la Figura 4.2 se grafican algunas funciones de distribución que ilustran las propiedades (i), (ii), y (iii). Observemos, en particular, que algunas de ellas pueden ser escalonadas, es decir, constantes por intervalos, mientras que otras son continuas, es decir, desprovistas de saltos.

Dado el conocimiento de la función de distribución, no solamente podemos obtener las probabilidades de $(T \leq x)$ para toda x , sino también las probabilidades de otros tipos de eventos. En particular tenemos

$$P(T > x) = 1 - F_T(x),$$

$$P(a < T \leq b) = F_T(b) - F_T(a),$$

y, en el caso de que los valores posibles de T son enteros,

$$P(T = n) = F_T(n) - F_T(n-1).$$

Como ejemplo, si suponemos conocida la función de distribución de T , entonces la probabilidad de que la componente falle después de 200 ciclos pero antes de 500 ciclos, sería $F_T(500) - F_T(200)$ mientras que la probabilidad de que dure más de 250 ciclos sería $1 - F_T(250)$. La probabilidad de durar exactamente 200 ciclos se calcularía como $F_T(200) - F_T(199)$. Debemos notar que si es conocida la función de distribución, el cálculo de las probabilidades anteriores no involucra la observación de un número grande de experimentos para luego calcular frecuencias relativas; las probabilidades mencionadas pueden obtenerse *a priori*.

El conocimiento de la función de distribución también nos permite calcular probabilidades condicionales. Supongamos, por ejemplo, que en un momento de inspección un componente ha acumulado 200 ciclos sin fallar y se desea cuantificar la posibilidad de que dure otros 100 ciclos adicionales. En otras palabras, se desea la probabilidad de que $T > 300$ dado que ya se observó que $T > 200$. La notación que se emplea para este concepto es $P(T > 300 | T > 200)$. En general, es posible demostrar que si un componente ha registrado M ciclos sin fallar, la probabilidad condicional de que dure otros K ciclos adicionales puede escribirse en términos de la función F_T como sigue:

$$P(T > M+K | T > M) = (1 - F_T(M+K)) / (1 - F_T(M)).$$

La interpretación de este número es la siguiente: Si fuese posible observar el desempeño del mismo componente un número grande de veces,

$P(T > M+K | T > M)$ representa la proporción aproximada de veces que el componente no fallaría antes de cumplir $M+K$ ciclos de uso, de entre las ocasiones en que haya durado por lo menos M ciclos. Esta interpretación tiene implicaciones en políticas de mantenimiento y reemplazo, pues si se decide dejar en su lugar un componente con M ciclos acumulados que no ha fallado, $P(T > M+K | T > M)$ representa una cuantificación de la *certeza* de que dure otros K ciclos adicionales. El valor de K puede representar, por ejemplo, el número de ciclos programados *antes de la siguiente inspección*. Para una cuantificación del *riesgo* que se toma al dejar en su lugar dicha refacción, puede calcularse

$$P(T \leq M+K | T > M) = 1 - P(T > M+K | T > M),$$

lo cual mide la probabilidad de que el componente con M ciclos acumulados que no ha fallado, lo haga antes de cumplir otros K ciclos adicionales. En el caso particular $K = 1$, la fórmula anterior proporciona la probabilidad de que un componente que no ha fallado durante M ciclos, lo haga durante *el siguiente ciclo de uso*. Este último concepto se incorpora en la función $R_T(M)$ definida para $M = 1, 2, \dots$ por

$$R_T(M) = P(T \leq M+1 | T > M) = (F_T(M+1) - F_T(M)) / (1 - F_T(M)),$$

la cual se conoce como la *Función de Riesgo* (o *Razón de Falla*) asociada a la variable aleatoria T . La función de riesgo es útil cuando se desea decidir el reemplazo de un solo componente disponible entre varias unidades de servicio.

Las funciones de distribución también pueden ser de utilidad para *comparar* distintos tipos de componentes. Supongamos que se tienen dos tipos o marcas de refacciones, A y B , y que ambas marcas están garantizadas por un periodo de 2000 ciclos de uso. Sea S el tiempo de falla para la componente de la marca A , y T el tiempo de falla para la marca B . Si $F_T(2000) < F_S(2000)$, es decir

$$P(T \leq 2000) < P(S \leq 2000),$$

entonces concluiríamos que el componente de marca A tiene mayor posibilidad de fallar antes de 2000 ciclos que la marca B . Es decir, la conclusión sería que la marca B es de mayor duración y/o calidad que la marca A , o bien que se recurriría con mayor frecuencia a la garantía cuando se utilizan refacciones de la marca A .

Otro tipo de problema que puede resolverse a través de la función de

distribución es el siguiente, referente a un programa de mantenimiento. Supongamos que se desea establecer un número fijo de ciclos de uso, C , para implementar la política de que la refacción sea reemplazada cuando se cumplan C ciclos de uso, aunque no haya fallado aún en ese momento. Si se elige un valor muy grande de C , las piezas tenderán a fallar antes de cumplir C ciclos y tendrán que ser reemplazadas solo después de haber manifestado una falla, lo cual puede ser muy costoso en algunos casos. Por otra parte, si a C se le asigna un valor muy pequeño, se estarán reemplazando en forma innecesaria muchas piezas que potencialmente aún mantenían vida útil. Un criterio para determinar el valor de C consiste en fijar la proporción *tolerable* de piezas que fallen antes de ser reemplazadas, denotada aquí por α , con $0 < \alpha < 1$. Sea C_α un número tal que $F_T(C_\alpha) = \alpha$. Entonces, si se opta por el reemplazo de piezas al cumplir éstas C_α ciclos de uso, solamente una proporción α de las piezas fallarán antes de su reemplazo programado. El valor C_α recibe el nombre de α -percentil de la función de distribución F_T .

Además de la función de distribución, existen otros criterios básicos asociados a una variable aleatoria T , que describen su comportamiento. Dos de los más usuales son el valor esperado, o esperanza, o media de T , denotado por $\mu(T)$, y la desviación estándar de T , denotada por $\sigma(T)$. Un tercer concepto, importante aunque menos difundido, es el Coeficiente de Sesgo, que aquí representaremos por $s(T)$. Las definiciones precisas de estas cantidades a partir de la función de distribución son también parte de un curso formal sobre probabilidad. En este reporte nos limitamos a mencionar las interpretaciones que pueden darse a estos tres números (en la Sección 3, se señalan las formas explícitas que tienen estas tres cantidades para los modelos particulares que se están considerando):

- La media $\mu(T)$ representa una medida de la localización o tendencia central de tiempos de falla. No obstante el nombre sugestivo "valor esperado", no debemos entender que un tiempo de falla T vaya a observarse con frecuencia en un valor exactamente igual a $\mu(T)$, sino que a la larga, los valores tenderán a ocurrir localizados alrededor de $\mu(T)$.

- La desviación estándar $\sigma(T)$, un número no-negativo, representa una medida de la dispersión que sufren los valores de T alrededor de su media $\mu(T)$. A mayor valor de la desviación estándar, los valores de T oscilan más alrededor de $\mu(T)$. A menor desviación estándar, los valores de T tienden a concentrarse más alrededor de $\mu(T)$. En efecto, la situación extrema

$\sigma(T) = 0$ significa que todos los valores de T resultan ser iguales a $\mu(T)$, es decir, T es determinístico con valor constante $\mu(T)$. El valor $\sigma^2(T)$ se conoce con el nombre de *varianza* de la variable aleatoria T .

— El coeficiente de sesgo $s(T)$ representa una medida de la *simetría* de la distribución de las distintas ocurrencias de T , y puede tomar cualquier valor (positivo, negativo, o cero). El valor $s(T) = 0$ corresponde a una distribución simétrica, es decir, cuando las mediciones de T tienden a ocurrir simétricamente alrededor de $\mu(T)$. Un valor $s(T) > 0$ corresponde a una variable T que tiene una tendencia mayor de ocurrencias ocasionales extremas por arriba de $\mu(T)$, y T se dice una variable *sesgada a la derecha*. Por el contrario, un valor $s(T) < 0$, refleja una mayor tendencia a que los valores extremos ocurran por abajo de $\mu(T)$, y T se dice *sesgada a la izquierda*.

En el ejemplo arriba mencionado sobre las dos marcas de refacciones, muy posiblemente encontraríamos $\mu(A) < \mu(B)$, lo cual refleja el hecho de que los componentes de marca A tienden a fallar antes que los de la marca B . En cuanto a las desviaciones estándar, supongamos que se cumple $\sigma(A) < \sigma(B)$. Esto significaría que a pesar de que su duración es estadísticamente más corta, los componentes de la marca A son más consistentes y predecibles en su desempeño, en el sentido de que la variabilidad de sus tiempos de vida alrededor de su media respectiva es menor.

Una relación universal entre $\mu(T)$ y $\sigma(T)$ es la que afirma la **Desigualdad de Chebyshev**:

$$P(|T - \mu(T)| \geq r) \leq \sigma^2(T)/r^2.$$

En otras palabras, la probabilidad de encontrar T fuera del intervalo $(\mu(T) - r, \mu(T) + r)$ es a lo más, $\sigma^2(T)/r^2$. Este resultado, de nueva cuenta, refleja el hecho de que a menor desviación estándar, T se concentra más alrededor de $\mu(T)$. Por ejemplo, si $\mu(T) = 20$ y $\sigma^2(T) = 9$, entonces la proporción de veces que T resulta con un valor fuera del intervalo $(16, 24)$ es a lo más, $9/16$, o 56%. Notemos que esta última conclusión se logra con el conocimiento único de los dos números $\mu(T)$ y $\sigma(T)$. Si además, la función de distribución fuese enteramente conocida, se podría calcular

$$P(|T - \mu(T)| \geq r) = 1 - \{F_T(\mu(T)+r) - F_T(\mu(T)-r)\}$$

sin tener que recurrir a la cota para la probabilidad que establece la desigualdad de Chebyshev.

3. EL MODELO BASADO EN UNA CADENA DE MARKOV

Hemos visto en la Sección 2, que conocida la función de distribución $F_T(x)$ para todos los valores de x , puede describirse el comportamiento aleatorio de los tiempos de falla, T . Sin embargo, en la práctica, la función F_T es desconocida. Supongamos que hemos tenido la experiencia de observar el desempeño de n componentes, y que los tiempos a los cuales éstos han fallado son T_1, T_2, \dots, T_n . La labor estadística consiste en utilizar estos valores de T para deducir F_T , es decir, estimar $F_T(x)$ para todos los valores de x .

En este trabajo consideraremos un modelo sencillo para daño acumulado, que puede visualizarse de la manera siguiente. Supondremos c estados de daño, s_1, s_2, \dots, s_c . De ellos, s_1 se identificará con la condición de ser una pieza recién instalada, mientras que s_c corresponderá a la condición de fallar después de usarse por un tiempo. Al progresar, cada estado representa mayor deterioro que el estado anterior. Durante su vida útil, la pieza o refacción va recorriendo los estados desde s_1 (cuando se instala) hasta s_c (cuando falla). Dicho recorrido se realiza, en efecto, de un modo aleatorio. Cuando se trata de daño acumulado, es natural suponer que una pieza no puede jamás pasar de un estado s_i a un estado s_j , con $j < i$.

Para modelar el ascenso aleatorio a través de los estados, se aplica el siguiente mecanismo. Supongamos que una pieza ya se encuentra en el estado de deterioro s_i , y que es sometida a un nuevo ciclo de uso. La acción de dicho ciclo de uso puede provocar que, con una probabilidad $q = 1-p$, la pieza pase del estado s_i al estado s_{i+1} , o bien, que con una probabilidad p , que la pieza permanezca en el estado s_i . La probabilidad q recibe el nombre de **probabilidad de transición** del estado i al estado $(i+1)$. Al contabilizar el número total de ciclos necesarios para llegar del estado s_1 al estado s_c , se obtiene el tiempo de falla T para esa pieza en particular. Debido al azar involucrado en pasar de un estado a otro con cada ciclo de uso, cada pieza recorre la cadena de estados a distintos tiempos; esto equivale a que cada pieza *falla* a distintos tiempos, lo cual es precisamente lo que pretendemos modelar. Notemos que la elección $q = 1$, lo cual afirma que cada ciclo de uso invariablemente causa un ascenso al siguiente estado de deterioro, produciría que T siempre fuese $c-1$, es decir, T se volvería una variable determinística. Por otra parte, el valor $q = 0$ corresponde a una pieza que nunca falla. Las aplicaciones realmente interesantes son para $0 < q < 1$.

Técnicamente hablando, el modelo descrito corresponde a una Cadena de Markov (ver Cap. 2 de [4]) con matriz de transición P, de dimensión cxc, dada por

$$P = \begin{bmatrix} p & q & 0 & 0 & \dots \\ 0 & p & q & 0 & \dots \\ \dots & & & & \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

y que comienza en el estado 1. La variable T corresponde al tiempo necesario para pasar del estado 1 de esta cadena al estado c. En el lenguaje propio de Cadenas de Markov, se dice que los estados 1, 2, ..., c-1 son transitivos si $q > 0$ mientras que el estado c es absorbente. A pesar de que existe una amplia teoría desarrollada acerca de las Cadenas de Markov, no mencionamos aquí todas sus propiedades puesto que el objeto de mayor interés por el momento es la función de distribución de T, la cual puede obtenerse directamente.

En efecto, es posible demostrar ([5]) que en el caso del modelo que se ha descrito, la función de distribución es

$$F_T(x) = \sum_{k=c-1}^{[x]} \binom{k-1}{k-c+1} p^{k-c+1} (1-p)^{c-1}, \text{ si } x \geq c-1, \text{ y}$$

$$F_T(x) = 0, \text{ si } x < c-1.$$

Aquí, la notación $[x]$ denota la parte entera de x, y

$$\binom{1}{m} = \frac{1!}{m!(1-m)!},$$

donde $m! = m(m-1)(m-2)\dots(2)(1)$.

La media y la desviación estándar de T también poseen expresiones en términos de c y p:

$$\mu(T) = (c-1)(1+p/(1-p)), \text{ y}$$

$$\sigma(T) = \sqrt{(c-1)(1+p/(1-p))(p/(1-p))}.$$

Con respecto al coeficiente de sesgo, es posible demostrar ([5]) que bajo las circunstancias descritas,

$$s(T) = (c-1)\{2-3(1-p)+(1-p)^2\} / \{(1-p)\sigma(T)\}^3.$$

Es importante señalar que $s(T) > 0$, excepto cuando $p = 0$ (en cuyo caso $s(T) = 0$). Esto significa que el modelo basado en cadenas de Markov no es adecuado para representar datos que muestren sesgo negativo. La restricción que esta condición ejerce sobre la aplicabilidad de estos modelos no es tan

grave como aparenta, puesto que el sesgo en la gran mayoría de los ejemplos reales de daño acumulado es no-negativo.

Con estas expresiones, el problema de estimar $F_T(x)$, $\mu(T)$, y $\sigma(T)$ se reduce a estimar los valores de c y p , los cuales pasan a constituir ahora los valores desconocidos. El problema, replanteado en estos términos, es encontrar estimaciones \hat{c} y \hat{p} , en base a las observaciones T_1, T_2, \dots, T_n . Debemos notar que con solamente *dos* números, c y p , se hace posible calcular, entre otras cosas, $F_T(x)$ para todo valor x (es decir, para un número infinito de x).

Un método de estimación se dice **consistente** si al crecer n , los valores \hat{c} y \hat{p} convergen (en algún sentido hecho preciso en teoría de probabilidad) a c y p , respectivamente. Esta propiedad significa que al coleccionar una muestra de tamaño mayor, se hace posible estimar con mayor precisión los verdaderos valores (desconocidos) de c y p . Sin embargo, en la práctica se cuenta únicamente con una muestra de tamaño n ; si n no es muy grande, es de esperarse que las estimaciones no serán necesariamente próximas a los valores reales desconocidos. Los métodos estadísticos tienen como uno de sus objetivos primordiales el cuantificar matemáticamente esta incertidumbre respecto a la precisión de las estimaciones y las inferencias que se efectúan posteriormente a partir de estos valores.

Volviendo al problema de estimar $F_T(x)$, debemos señalar que existe una forma muy intuitiva de lograrlo, mediante la llamada **Función de Distribución Empírica**,

$$\tilde{F}_T(x) = \frac{\text{número de observaciones } \leq x}{n}$$

La intuición detrás de dicho estimador es muy clara: La estimación para $P(T \leq x)$ es simplemente la *proporción observada* de valores de T_1, T_2, \dots, T_n que cumplen ser menores o iguales a x . Las funciones de distribución empíricas son *escalonadas*, es decir, son constantes por intervalos y tienen saltos solamente en los valores de x que coincidan con algún valor observado de T (ver Figura 4.2). De hecho, la función \tilde{F}_T podría reemplazarse por la verdadera F_T en las expresiones para probabilidades de la Sección 2, para obtener una versión estimada de las mismas. Por otra parte, si adoptáramos el modelo anteriormente descrito y obtuviéramos estimaciones \hat{c} y \hat{p} , contaríamos una estimación alterna para F_T dada por

$$\hat{F}_T(x) = \sum_{k=\hat{c}-1}^{\lfloor x \rfloor} \binom{k-1}{k-\hat{c}+1} p^{\hat{c}-k+1} (1-p)^{\hat{c}-1}, \text{ si } x \geq \hat{c}-1, \text{ y}$$

$$\hat{F}_T(x) = 0, \text{ si } x < \hat{c}-1.$$

Así mismo, por substitución, calculamos estimaciones de $\mu(T)$ y $\sigma^2(T)$ dadas por

$$\hat{\mu}(T) = (\hat{c}-1)(1+p/(1-p)), \text{ y}$$

$$\hat{\sigma}^2(T) = (\hat{c}-1)(1+p/(1-p))(p/(1-p)).$$

El estimador \hat{F}_T es un ejemplo de un estimador paramétrico (los parámetros son c y p) mientras que el estimador \tilde{F}_T es un estimador no-paramétrico. En general, la teoría estadística establece que si las condiciones del modelo postulado son satisfechas, entonces los estimadores paramétricos son más eficientes que los no-paramétricos. Esto significa que la convergencia de $\hat{F}_T(x)$ a $F_T(x)$ ocurre con mayor rapidez que la de $\tilde{F}_T(x)$. En otras palabras, a un mismo tamaño de muestra n , $\hat{F}_T(x)$ tiene mayor *precisión* que $\tilde{F}_T(x)$. Es por esta razón que el estimador paramétrico \hat{F}_T aquí descrito es preferible sobre \tilde{F}_T .

Con lo anterior, hemos reducido el problema de estimación de F_T a encontrar estimadores consistentes de c y p , en base a las n observaciones. Mencionaremos los resultados de dos métodos que se han propuesto para ello, llamados el Método de Momentos (MM) y el Método de la Función Generatriz de Probabilidades (MFGP). Para una descripción precisa acerca de estos dos métodos y cómo dependen de las observaciones, debe consultarse la referencia [5].

El MM es un método clásico para efectuar estimación en estadística, el cual se fundamenta en el cálculo de ciertos promedios. Para el caso específico de modelos de daño de acumulado, su descripción se encuentra en [5] y [2].

El MFGP es un procedimiento nuevo, propuesto en [5], que se basa en las propiedades de ciertas funciones de la teoría de probabilidad aplicada a variables aleatorias que toman solo valores enteros. Comparado al MM, el cálculo de los estimadores con el MFGP es más complejo, pero se ha demostrado ([5]) que el MFGP puede ser mejor que el MM en muchos casos. Más precisamente, podemos mencionar que mientras ambos métodos son consistentes para c y p , la evidencia muestra que el MFGP puede ser más eficiente que el MM en muchos casos. En la Sección 4, describimos el empleo de un programa de

cómputo que ha sido desarrollado para calcular tanto estos dos estimadores como las funciones de distribución estimadas. En lo subsecuente, utilizaremos la notación \hat{p} y \hat{c} para las estimaciones de p y c , obtenidas en base a las observaciones T_1, T_2, \dots, T_n por ambos métodos. Como mencionamos anteriormente, estos valores automáticamente proporcionan estimaciones $\hat{F}_T(x)$ para toda x , $\hat{\mu}(T)$, $\hat{\sigma}^2(T)$, y $\hat{s}(T)$.

Cabe notar que una vez que se tiene disponible un método para estimar c y p , no es necesario identificar los c estados de deterioro explícitamente. La variable de interés es el tiempo que transcurre entre la instalación y el retiro de la pieza, sin importar lo que ocurra entre estos dos eventos. De hecho, los c estados que se postulan y la probabilidad de transición q entre ellos son sólo un artificio para describir explícitamente la función de distribución $F_T(x)$. Una posible interpretación es que c representa la *longitud* del recorrido entre instalación y falla, mientras que q tiene que ver con la *velocidad* con la cual se recorre esta distancia. La condición de que q sea constante en cada estado de deterioro, refleja la suposición de que no existen cambios temporales en el mecanismo de acumulación de daño, en las propiedades de los materiales, ni en las condiciones de trabajo a la que es sometida la pieza. Desde luego, habrá ejemplos no cubiertos por el presente trabajo, en los cuales la velocidad de la acumulación del daño se incremente al transcurrir el tiempo y el estado de daño, es decir,

$$0 < q_1 < q_2 < \dots < q_{c-1} < 1$$

donde q_i es la probabilidad de pasar del estado i al estado $(i+1)$.

4. USO DEL PROGRAMA DE COMPUTO

4.1. Antecedentes

El programa de cómputo toma por entrada una serie T_1, T_2, \dots, T_n de tiempos de falla observados en ciertas componentes y calcula las estimaciones \hat{p} , \hat{q} , \hat{c} , $\hat{F}_T(x)$, $\hat{\mu}(T)$, $\hat{\sigma}^2(T)$, y $\hat{s}(T)$, utilizando ambos métodos señalados en la Sección 3 (MM y MFGP). Las funciones de distribución estimadas se grafican en la pantalla. Dada una colección de tiempos de falla, el programa es automático; corresponde al usuario verificar que la aplicación pretendida realmente se ajuste a las suposiciones del modelo:

1. Los tiempos T_1, T_2, \dots, T_n han sido medidos en unidades que son *ciclos de uso*.
2. Los n tiempos han sido obtenidos con el mismo tipo de

componentes y bajo las mismas circunstancias de uso, o al menos en componentes para los cuales sí puede suponerse una *misma* función de distribución para sus vidas útiles.

3. Las componentes fueron instaladas *nuevas*.

4. Los tiempos de fallas entre componentes son **independientes**, es decir, la actuación de alguno ellos no transmite peculiaridades al desempeño de otro dentro de la colección.

Un gran número de situaciones reales sí serán cubiertas por estas condiciones. Si alguna de las suposiciones no es satisfecha, las propiedades teóricas de los estimadores – consistencia y eficiencia, por ejemplo – no necesariamente se cumplen, de modo que la función de distribución estimada que reporta el programa puede ser inadecuada. Este modelo, como cualquier otro modelo estadístico, puede ser potencialmente útil cuando se utiliza prudentemente; no debe esperarse que toda situación de daño acumulado sea cubierta por el tipo particular de análisis descrito en el presente trabajo.

4.2 Manejo del programa.

El programa **DAMAGE**, disponible con los autores de el presente trabajo, está escrito en Turbo C (V 2.0), y consiste de rutinas numéricas, de estadística y de graficación. La estadística contiene los dos métodos ya mencionados de estimación de los parámetros. La porción de graficación contiene subrutinas para el ambiente gráfico de la máquina donde se ejecute y para graficar las funciones de distribución tanto estimadas como la empírica.

4.2.1 Ejecucion del programa DAMAGE

Se presupone que se cuenta con un archivo ASCII que contiene los datos observados, llamado por ejemplo DATOS.DAT. Dicho archivo de datos puede ser creado con cualquier editor, consta de $(n+3)$ líneas, y debe estar conformado como sigue:

Primera línea:	un título para el caso
Segunda línea:	el número "1"
Tercera línea:	un nombre corto para la variable que se estudia
Cuarta línea:	dato 1 (T_1)
Quinta línea:	dato 2 (T_2)
...	
última línea:	dato n (T_n)

Si en un futuro se obtienen observaciones adicionales y se desean incluir en el análisis, bastaría agregar las nuevas observaciones al archivo DATOS.DAT, sin modificar ninguna de las líneas previamente existentes.

Suponiendo que el disco del programa y el archivo DATOS.DAT se encuentran localizados en el *drive* A, el programa se ejecuta de la siguiente manera:

```
A:\>DAMAGE DATOS.DAT <enter>.
```

Después de la lectura correspondiente de los datos, el programa procede a calcular las estimaciones y a evaluar las funciones de distribución. Dependiendo del tamaño de la muestra y del problema, esto puede tomar algunos minutos. Cuando se concluyen los cálculos necesarios, se muestra en la pantalla el menú principal. Las opciones que se ofrecen son entonces:

1) Estadísticas básicas

Produce los resultados de las estimaciones de los parámetros por ambos métodos (MM y MFGP), medias, varianzas, y coeficientes de sesgo.

2) Graficación de funciones de distribución estimadas

Produce una gráfica en la pantalla de las funciones de distribución estimadas, así como la función de distribución empírica.

3) Tabla de percentiles

Produce los α -percentiles C_a , para valores de α preseleccionados.

4) Tabla de la función de distribución

Produce una tabulación de la función de distribución $\hat{F}_T(x)$, para valores distintos de x .

5) Cálculo de cuantiles arbitrarios

Permite el cálculo de percentiles C_α para valores de α no incluidos en la opción 3.

6) Fin

Terminación del programa DAMAGE.

4.2.2 Ejemplo

Presentamos aquí un ejemplo aplicado a un juego de datos reportado en [2], pág. 183. Las observaciones corresponden a mediciones de fatiga en 20 muestras de tiras ranuradas de metal de 5mm de grosor. Los datos fueron medidos en número de ciclos de uso hasta fatiga, y son los siguientes:

100, 110, 133, 135, 140, 157, 163, 165, 165, 170,
180, 183, 185, 200, 201, 205, 215, 240, 253, 280.

El archivo de datos para aplicar el programa DAMAGE a este ejemplo está formado por las siguientes líneas:

```
FATIGA EN MUESTRAS DE METAL
1
CICLOS
100
110
.
. etc.
.
253
280
```

La ejecución del programa DAMAGE produce en sus opciones 1, 2, 3, y 4 las Figuras 4.1, 4.2, 4.3, y 4.4 respectivamente.

Fig. 4.1: Se observa que la media estimada de los datos es 179 y la varianza es 2112.421. El coeficiente de sesgo es 0.341 (positivo), de manera que el modelo es factible (ver Sección 2).

Fig. 4.2: Se observa cómo las funciones de distribución estimadas por los dos métodos pueden ser muy diferentes entre sí, aunque ambas se ajustan a la función de distribución empírica en forma excelente.

Fig. 4.3: Se observa, por ejemplo, para $\alpha = .05$, $C_\alpha = 107$. Esto significa que 5% de las piezas de metal sufren fatiga antes de 107 ciclos de uso. Si se desea C_α para un valor de α no considerado en esta tabla, deberá recurrirse a la Opción 5.

Fig. 4.4: Como ejemplo, para $x = 140$, puede leerse $\hat{F}_T(x) = 0.231$ (23.1% de las piezas sufren fatiga antes de 140 ciclos de uso). Con las lecturas de \hat{F}_T obtenidas de esta tabla, pueden calcularse las probabilidades condicionales, así como las funciones de riesgo descritas en la Sección 2.

5. BIBLIOGRAFIA

- [1] Bhattacharyya, G. K. and Johnson, R. A. (1977) "Statistical Concepts and Methods", John Wiley & Sons.
- [2] Bogdanoff, J. L., and Kozin, F. (1985) "Probabilistic Models of Cumulative Damage", John Wiley & Sons.
- [3] Bowker, A. H. y Liéberman, G. J. (1981) "Estadística para Ingenieros", Prentice-Hall Hispanoamericana.
- [4] Papoulis, A. (1984) "Probability, Random Variables, and Stochastic Processes", McGraw-Hill.
- [5] Pérez, F. J. (1991) "Estimación en los Modelos de Daño Acumulado Basados en Cadenas de Markov", Tesis Profesional, Facultad de Ciencias Físico-Matemáticas, Universidad Autónoma de Nuevo León.

```

*****
*
*           ** MENU DE OPCIONES **
*
*   (1) Estadisticas y Estimaciones
*   (2) Graficas de F. de Dist.
*   (3) Tabla de Cuantiles
*   (4) Tabla de F.D
*   (5) Calcula un cuantil
*   (6) Salir
*
*   Opcion -->
*
*
*****

```

```

*****
*
*           PARAMETROS
*
*   MOMENTOS          FGP
* ESTADOS      14          15
* PROBAB.    0.922        0.920
*
*
*           ESTADISTICAS
*
*   MUESTRA    TEORICAS
* MEDIA  179.0000  174.819
* VAR    2112.421  2008.158
* SESGO  0.34122   0.53499
*
* NUMERO DE DATOS = 20
* VALOR MINIMO    = 110
* VALOR MAXIMO    = 280
*
* Presione <ENTER> para ir al menu...
*
*****

```

FIGURA 4.1

FATIGAS EN MUESTRAS DE METAL

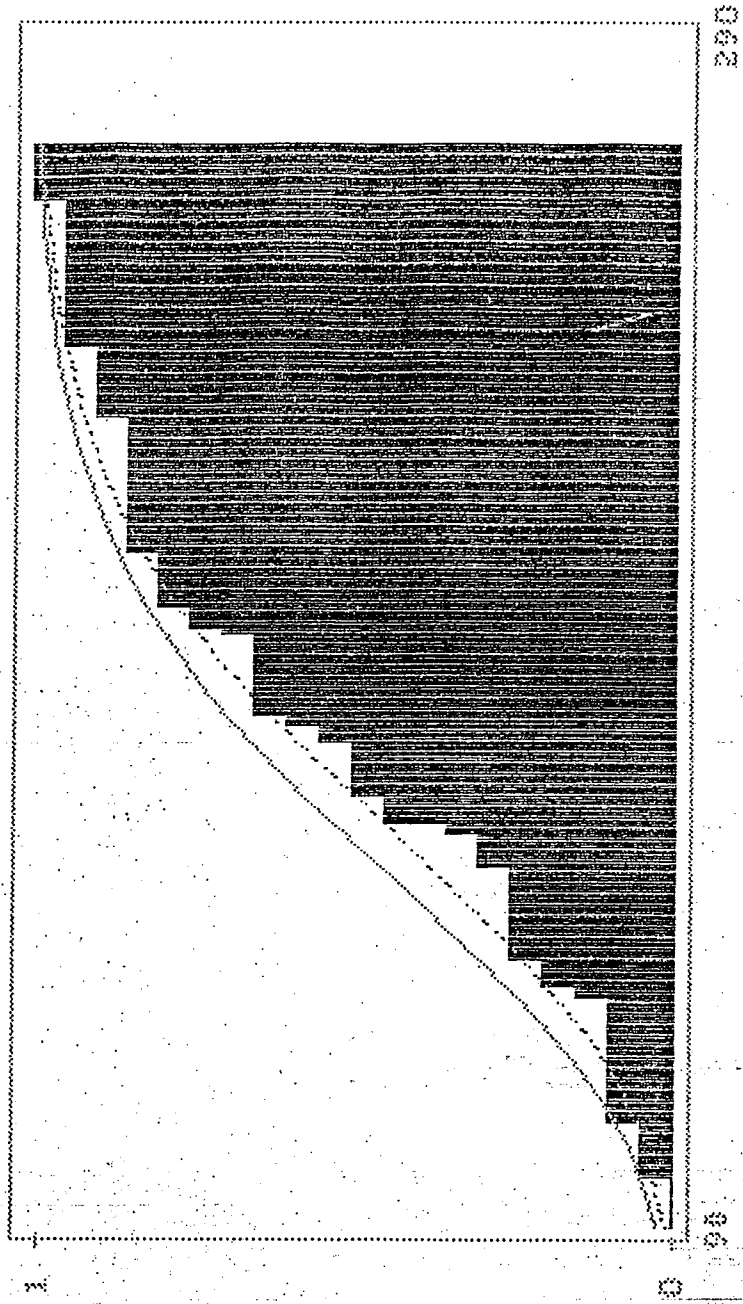


FIGURA 4.2

FD MOMENTOS FD FCP

```

*****
CUANTILES
T      F(T)
107 , 0.050000
122 , 0.100000
142 , 0.250000
172 , 0.500000
202 , 0.750000
232 , 0.900000
*****

```

Presione ENTER para ir al menu...

FIGURA 4.3

VALORES DE LA FUNCION DE DISTRIBUCION (FGP)

```

*****
T      F(T)
100 , 0.028453
110 , 0.056682
120 , 0.095674
130 , 0.158336
140 , 0.231316
150 , 0.315250
160 , 0.405468
170 , 0.496902
180 , 0.584877
190 , 0.665726
200 , 0.737064
210 , 0.797747
220 , 0.847678
230 , 0.887551
240 , 0.918535
250 , 0.942020
260 , 0.959421
270 , 0.972044
280 , 0.981022
290 , 0.987294
*****

```

Presione ENTER para salir al menu...

FIGURA 4.4