*Revista Colombiana de Estadística*
*Nºs 19-20, junio-diciembre 1989*

# UN ESTADISTICO DE PRUEBA PARA LA HIPOTESIS DE CONTAGIO DISTRIBUIDO EN UN PROCESO DE ACCIDENTES

*Antonio Velasco Muñoz* - *Freddy Angel Amaya Robayo*

Profesor Asociado, Universidad Nacional

Ms. en Estadística, Universidad Nacional

## 1. Antededentes.

Cuando se modela el número de accidentes sufridos por un individuo en un intervalo de tiempo, se piensa generalmente en una distribución de Poissón la cual se determina por un parámetro $\lambda$ que representa la media teórica de accidentes por unidad de tiempo, el cual permanece constante durante el tiempo de observación y si no registra la influencia, sobre la probabilidad de la ocurrencia de accidentes pasados.

Se considera que cada accidente pueda alterar la propensión a más accidentes y se obtendrá una distribución de probabilidad más general, que dependerá de los parámetros $\lambda_o, \lambda_1, \lambda_2, \ldots \lambda_m$, donde $\lambda_i$ representa la propensión a accidentes cuando se han sufrido $i$-accidentes.

Si $\lambda_o = \lambda_1 = \ldots = \lambda_m = \ldots = \lambda$ se obtiene entonces la forma de la distribución de Poissón.

Greenwood y Youle (1920) propusieron el primer modelo de probabilidad el cual conduce a una distribución de Poisson. Posteriormente Polya (1930) propone un nuevo modelo usando los siguientes postulados:

PI - Todos los individuos presentan igual probabilidad de sufrir un accidente (Se denotó como No-MEZCLA).

PII - Es posible que existe contagio por accidentes.

PIII - La experiencia al sufir accidentes influyen en el individuo y hacen modificar la probabilidad de que sufra accidentes futuros (EFECTO DE LA EXPERIENCIA EN EL TIEMPO).

Estos postulados conocidos como el Esquema de Polya fueron estudiados por Bates y Newman (1952) quienes presentan un modelo general que responde a la probabilidad de que un individuo sufra un accidente FATAL y la probabilidad de supervivencia después de un accidente. Estos autores examinan la hipótesis de ausencia de contagio utilizando la distribución del número de accidentes. Sin embargo la pruba no tiene una potencia satisfactoria. Con el uso de las distribuciones de tiempo transcurrido desde el inicio de la observación hasta la ocurrencia del $i$-ésimo accidente, Bates (1955) deduce una prueba de potencia uniforme máxima (P.U.M.) para la hipótesis de ausencia de contagio versus contagio positivo (más propenso de sufrir accidentes) versus contagio negativo (los accidentes previos "le enseñan" a evitar nuevos accidentes). En 1972 Neyman propone estudiar la aleatoridad del parámetro de contagio posiblemente mediante los test $C(\alpha)$. Se presentan resultados en esta dirección y su aplicación a información del Ingenio Risaralda durante 1984.

**Objetivo.** Es proponer una estadística de prueba para examinar aleatoridad en el contagio por accidentes bajo el supuesto del Esquema de Polya.

## 2. Modelo probabilístico.

Se considera que un individuo $I$ está expuesto desde el momento $t = 0$ a un accidente FATAL.

Sea $P_{m,n}(T_1,T_2)$ la probabilidad de que dicho individuo sufra $n$ accidentes durante el intervalo $(T_1,T_2)$ dado que ha sufrido hasta $T_1$, $m$ accidentes.

El resultado de Polya se resume así:

**T1:** Si $T_2$ tiende a $T_1$ entonces la probabilidad $P_{m,n}(T_1,T_2)$ tiende a $P_{m,n}(T_1,T_1)$ lo que implica que

$$P_{mn}(T_1,T_1) = \begin{cases} 1 & \text{si} \quad n = 0 \\ 0 & \text{si} \quad n > 0 \,. \end{cases}$$

**T2:** La probabilidad $P_{m,n}(T_1,T_2)$ depende del número previo de accidentes $m$ y del valor $T_1$ pero no de los momentos en que estos accidentes ocurrieron.

**T3:** Si $T_2 = T_1$ la probabilidad $P_{m,n}(T_1,T_2)$ es derivable respecto a $T_2$ y además:

$$\frac{\partial}{\partial T_2} P_{m,n}(T_1,T_2)\Big|_{T_2=T_1} \begin{cases} \dfrac{-\lambda_m}{1+\upsilon T_1} & \text{si} \quad n = 0 \\ \dfrac{+\lambda_m}{1+\upsilon T_1} & \text{si} \quad n = 1 \\ 0 & \text{si} \quad n > 1 \end{cases}$$

Los valores $\lambda_o, \lambda_1, \ldots, \lambda_m, \ldots$ son positivos y se denominan parámetros de contagio; $\upsilon > 0$ se denomina parámetro de efecto del tiempo, o experiencia.

El contagio por accidentes está determinado por la relación entre los $\lambda$ :

Si $\lambda_o < \lambda_1 < \lambda_2 < \ldots < \lambda_m < \ldots$ el contagio se denomina positivo.

Si $\lambda_o > \lambda_1 > \lambda_2 > \ldots > \lambda_m > \ldots$ el contagio se denomina negativo

definiendo $\psi_{m+k} = \lambda_{m+k+1} - \lambda_{m+k}$, entonces, el contagio es positivo si $\psi_{m+k} > 0$, negativo si $\psi_{m+k} < 0$ y no existe si $\psi_{m+k} = 0$ y el contagio por accidentes puede estudiarse en base a $\psi_{m+k}$. De acuerdo con la forma de los parámetros $\psi_i$ y $\upsilon$ se pueden considerar varios modelos probabilísticos:

i) $\psi_i = 0$ para todo $i$, $\upsilon > 0$; denominado no mezcla - posible efecto del tiempo-contagio nulo (no contagio).

ii) $\psi_m, \psi_{m+1}, \ldots, \psi_{m+n-1}$ posiblemente diferentes; $\upsilon_o = 0$ no mezcla-ausencia de efecto del tiempo - posible contagio.

iii) $\psi_m = \psi_{m+1} = \psi_{m+2} = \ldots = \psi_{m+n-1} = \psi$; $\upsilon > 0$ no mezcla - posible efecto del tiempo - contagio lineal.

iv) $\psi_m = \psi_{m+1} = \ldots = \psi_{m+n-1} = \psi$; $\upsilon = 0$ no mezcla - ausencia de efecto del tiempo - contagio lineal.

v) $\psi_m = \psi_{m+1} = \ldots = \psi_{m+n-1} = 0$ ; $\quad \nu = 0$ no mezcla - ausencia de efecto del tiempo - no contagio.

La atención se centra en los modelos (iii) y (iv).

Las distribuciones de probabilidad $P_{m,n}(T_1,T_2)$ están dadas por las fórmulas:

$$P_{m,o}(T_1,T_2) = \left(\frac{1+\nu T_1}{1+\nu T_2}\right)^{(\lambda_m/\nu)}$$

$$P_{m,n}(T_1,T_2) = (-1)^n \lambda_m \lambda_{m+1} \cdots \lambda_{m+n-1} \sum_{k=o}^{n} \left(\frac{1+\nu T_1}{1+\nu T_2}\right)^{(\lambda_m+k)/\nu} D_{kn}^{-1}$$

donde

$$D_{k,n} = \prod_{\substack{j=o \\ j\neq k}}^{n} (\lambda_{m+k} - \lambda_{m+j}) \; ; \qquad n \geqslant 1.$$

De las fórmulas dadas anteriormente se puede deducir la función de densidad de probabilidad de las variables $\tau_i$, $i = 1,2,\ldots,n$ donde $\tau_i$ es el tiempo transcurrido desde el inicio de la observación hasta la ocurrencia del $i$-ésimo accidente, suponiendo que el intervalo de observación es unitario, es decir $(T_1,T_1+1)$ tal función es:

$$f_{\tau_1,\tau_2,\ldots,\tau_n}(t_1,t_2,\ldots,t_n \mid \psi_m,\ldots,\psi_{m+n-1};\ \nu)$$

$$= \frac{\prod_{k=1}^{n} (1+\nu t_k/a)^{((\psi_{m+k})/\nu)-1}}{a^n \sum_{k=o}^{n} (-1)^k (1+\nu t_k/a)^{(\psi_{m+k}+\psi_{m+k+1}+\cdots+\psi_{m+n-1})/\nu} R_{k,n}^{-1}}$$

$$0 < t_1 < t_2 < \ldots < t_n < 1 \qquad a = 1+\nu T_1$$

$$\frac{\partial}{\partial T_2} P_{m,n}(T_1,T_2)\Big|_{T_2=T_1} \begin{cases} \dfrac{-\lambda_m}{1+\nu T_1} & \text{si} \quad n = 0 \\ \dfrac{+\lambda_m}{1+\nu T_1} & \text{si} \quad n = 1 \\ 0 & \text{si} \quad n > 1 \end{cases}$$

Los valores $\lambda_o, \lambda_1, \ldots, \lambda_m, \ldots$ son positivos y se denominan parámetros de contagio; $\nu > 0$ se denomina parámetro de efecto del tiempo, o experiencia.

El contagio por accidentes está determinado por la relación entre los $\lambda$ :

Si $\lambda_o < \lambda_1 < \lambda_2 < \ldots < \lambda_m < \ldots$ el contagio se denomina positivo.

Si $\lambda_o > \lambda_1 > \lambda_2 > \ldots > \lambda_m > \ldots$ el contagio se denomina negativo

definiendo $\psi_{m+k} = \lambda_{m+k+1} - \lambda_{m+k}$, entonces, el contagio es positivo si $\psi_{m+k} > 0$, negativo si $\psi_{m+k} < 0$ y no existe si $\psi_{m+k} = 0$ y el contagio por accidentes puede estudiarse en base a $\psi_{m+k}$. De acuerdo con la forma de los parámetros $\psi_i$ y $\nu$ se pueden considerar varios modelos probabilísticos:

i) $\psi_i = 0$ para todo $i$, $\nu > 0$; denominado no mezcla - posible efecto del tiempo-contagio nulo (no contagio).

ii) $\psi_m, \psi_{m+1}, \ldots, \psi_{m+n-1}$ posiblemente diferentes; $\nu_o = 0$ no mezcla-ausencia de efecto del tiempo - posible contagio.

iii) $\psi_m = \psi_{m+1} = \psi_{m+2} = \ldots = \psi_{m+n-1} = \psi$; $\nu > 0$ no mezcla - posible efecto del tiempo - contagio lineal.

iv) $\psi_m = \psi_{m+1} = \ldots = \psi_{m+n-1} = \psi$; $\nu = 0$ no mezcla - ausencia de efecto del tiempo - contagio lineal.

v) $\psi_m = \psi_{m+1} = \ldots = \psi_{m+n-1} = 0$ ; $\nu = 0$ no mezcla - ausencia de efecto del tiempo - no contagio.

La atención se centra en los modelos (iii) y (iv).

Las distribuciones de probabilidad $P_{m,n}(T_1,T_2)$ están dadas por las fórmulas:

$$P_{m,o}(T_1,T_2) = \left(\frac{1+\nu T_1}{1+\nu T_2}\right)^{(\lambda_m/\nu)}$$

$$P_{m,n}(T_1,T_2) = (-1)^n \lambda_m \lambda_{m+1} \cdots \lambda_{m+n-1} \sum_{k=o}^{n} \left(\frac{1+\nu T_1}{1+\nu T_2}\right)^{(\lambda_m+k)/\nu} D_{kn}^{-1}$$

donde

$$D_{k,n} = \prod_{\substack{j=o \\ j\neq k}}^{n} (\lambda_{m+k} - \lambda_{m+j}) \; ; \qquad n \geqslant 1.$$

De las fórmulas dadas anteriormente se puede deducir la función de densidad de probabilidad de las variables $\tau_i$, $i = 1,2,\ldots,n$ donde $\tau_i$ es el tiempo transcurrido desde el inicio de la observación hasta la ocurrencia del $i$-ésimo accidente, suponiendo que el intervalo de observación es unitario, es decir $(T_1,T_1+1)$ tal función es:

$$f_{\tau_1,\tau_2,\ldots,\tau_n}(t_1,t_2,\ldots,t_n \mid \psi_m,\ldots,\psi_{m+n-1}; \nu)$$

$$= \frac{\prod_{k=1}^{n} (1+\nu t_k/a)^{((\psi_{m+k})/\nu)-1}}{a^n \sum_{k=o}^{n} (-1)^k (1+\nu t_{k}/a)^{(\psi_{m+k}+\psi_{m+k+1}+\cdots+\psi_{m+n-1})/\nu} R_{k,n}^{-1}}$$

$$0 < t_1 < t_2 < \ldots < t_n < 1 \qquad a = 1+\nu T_1$$

$$R_{k,n} = (\psi_m+\psi_{m+1}+\ldots+\psi_{m+k-1})(\psi_{m+1}+\psi_{m+2}+\ldots+\psi_{m+n-1})\ldots$$
$$(\psi_{m+k-2}+\psi_{m+k-1})\psi_{m+k-1}\psi_{m+k}(\psi_{m+k}+\psi_{m+k+1})\ldots$$
$$(\psi_{m+k}+\psi_{m+k+1}+\ldots+\psi_{m+n-1}).$$

De esta fórmula, tomando los límites necesarios, se establecen las funciones de densidad para los casos que interesan y que se denominan modelo I y II:

**I)** No mezcla - ausencia de efecto del tiempo - contagio lineal

$$f_{\tau_1,\ldots,\tau_n}(t_1,t_2,\ldots,t_n|\psi_i=\psi\,;\,\nu=0) = n!\left(\frac{\psi}{e^{\psi}-1}\right)^n e^{\psi\sum_{k=1}^{n} t_k}$$

**II)** No mezcla - posible efecto del tiempo - contagio lineal

$$f_{\tau_1,\ldots,\tau_n}(t_1,\ldots,t_n|\psi_i=\psi\,;\,\nu) = \frac{n!\;\psi^n}{a^n\left[(1+\nu/a)^{\psi/\nu}\right]^n}\prod_{k=1}^{n}\left(1+\frac{\nu t_k}{a}\right)^{(\psi/\nu)-1}$$

$$0 < t_1 < t_2 < \ldots < t_n < 1 .$$

Considerando como momento inicial $T_1 = 0$, y que las distribuciones anteriores pueden verse como distribución de estadísticos de orden de una muestra aleatoria de la variable aleatoria $\tau$ con función de densidad de probabilidad.

$$f_\tau(t|\psi) = \frac{\psi}{e^{\psi}-1}\, e^{\psi t} \qquad 0 < t < 1$$

$$f_\tau(t|\psi,\nu) = \frac{\psi}{(1+\nu)^{\psi/\nu}-1}\,(1+\nu t)^{(\psi/\nu)-1} \qquad 0 < t < 1 .$$

El problema se reduce, considerablemente, a tratar con funciones más sencillas, facilitándose así la estimación de parámetros.

## 3. Los test $C(\alpha)$.

Antes de continuar con el análisis de cada modelo, se esboza el método a seguir en la deducción del estadístico de prueba para la hipótesis de aleatoriedad del contagio; este método se denomina, método de los **Test** $C(\alpha)$ descrito así:

Dada una variable aleatoria $X$ con valores en $\mathcal{X}$, cuya función de densidad depende de dos parámetros $\xi$ escalar y $\theta$ vector $\theta = (\theta_1,\theta_2,\ldots,\theta_s)$, $s \geqslant 1$; se nota dicha función por $f(x|\xi,\theta)$ indicamos la dependencia de la función $f$ y su distribución respecto a $\xi$ y $\theta$ por $X(\xi,\theta)$; consideremos las hipótesis:

$$H_o : \xi = 0 \quad \text{vs} \quad H_1 : \xi \neq 0$$

permanece desconocido.

El método se inicial seleccionando una función $h$ de valor real y dominio $\mathcal{X}$, de modo que la variable aleatoria $h[X(\xi,\theta)]$ tengan media y varianza finitas, $h(\xi,\theta)$, $\sigma^2(\xi,\theta)$ respectivamente. Luego se construye una función provisional de prueba $g^*(x,\theta) = h(x)-h(\xi,\theta)$ definida en $\mathcal{X}\times\Theta$, $g^*[X(0,\theta),\theta]$ tiene media cero y varianza $\sigma^2(0,\theta)$, luego si $X_1,X_2,\ldots,X_n$ en una muestra aleatoria de $X$ y $H_o$ es verdadera

$$Z_n(\theta) = \frac{\sum_{i=1}^{n} g^*[X_i,\theta]}{\sqrt{n}\,\sigma(0,\theta)} \longrightarrow N(0,1) \;,$$

sin embargo $Z_n^*(\theta)$ no se puede utilizar como estadístico de prueba puesto que no se puede calcular de las observaciones por depender de $\theta$ que es desconocido; se puede pensar en estimar $\theta$ pero $Z_n^*(\hat{\theta})$ no es, necesariamente, normal; Neyman 1972 demuestra que para que $Z_n^*(\theta)-Z_n^*(\hat{\theta})$ tienda a cero en probabili dad es necesario y suficiente que la función provisional de prueba sea ortogonal a las derivadas:

$$y_j = \frac{\partial LN f}{\partial \theta_j}\Big|_{\xi=0} \qquad j = 1,2,\ldots,s.$$

Las condiciones de ortogonalidad se satisfacen si reemplazamos $g^*$ por $g = g^* - \sum_{j=1}^{s} a_j y_j$.

$a_j$ son los coeficientes de la regresión de $g^*$ sobre $y_1, \ldots, y_s$ así el estadístico de prueba sería:

$$Z_n(\hat{\theta}) = \frac{\sum_{i=1}^{n} g[X_i,\hat{\theta}]}{\{nV(g)\}^{\frac{1}{2}}} .$$

La función de potencia está dada por

$$B(\xi,\theta) = 1 - \int_{-z_{\alpha/2}}^{z_{\alpha/2}} \frac{e^{-\frac{1}{2}(x-\omega)^2}}{\sqrt{2\pi}}\, dx$$

$\omega = \xi\sqrt{n}\sigma_h\rho$ es el parámetro de no centralidad, donde $h = \frac{\partial LN f}{\partial \xi}\Big|_{\xi=0}$, para definir $\rho$ y $\sigma_h$ construimos $h - \sum a_j^o y_j$, los $a_j^o$ son coeficientes de la regresión de $h$ sobre $y_j$ calculados asumiendo $\xi = 0$; $\sigma_h$ es la desviación de

$$h[X(0,\theta),\theta] - \sum a_j^o\, y_j[X(0,\theta),\theta]$$

$\rho$ es el coeficiente de correlación entre $g$ y la expresión anterior, la cual es la función óptima de prueba.

# 4. Estimación del parámetro $\psi$ modelo $I$.

## 4.1. METODO DE MAXIMA VEROSIMILITUD.

Sea $t_1, t_2, \ldots, t_n$ muestra aleatoria de $\tau$, la función de verosimilitud esta dada por:

$$L(\psi) = \left(\frac{\psi}{e^{\psi}-1}\right)^n e^{\psi \sum_{i=1}^{n} t_i}$$

tomando logaritmo y derivando se obtiene la ecuación de verosimilitud

$$\frac{e^{\psi}-\psi e^{\psi}-1}{\psi(e^{\psi}-1)} + \bar{t} = 0$$

la cual se puede escribir como: $e^{\psi}+\psi e^{\psi}(\bar{t}-1)-\psi\bar{t}-1=0$ ecuación que no se puede resolver algebraicamente. Aproximando por polinomios de Taylor el término $e^{\psi}$ y resolviendo la ecuación resultante obtenemos el estimador:

$$\hat{\psi} = \frac{1-2\bar{t}}{\bar{t}-1}$$

Al medir mediante simulación el error medio cuadrático de este estimador, resuelta demasiado alto, casi igual al rango, lo que descarta su utilidad, se procede entonces por el método de los momentos.

## 4.2. METODO DE LOS MOMENTOS.

Por este método se trata de resolver el sistema de ecuaciones:

$$\left.\begin{aligned} E(\tau) &= \bar{t} \\ V(\tau) &= S^2 \end{aligned}\right\} \quad \text{obteniéndose:} \quad e^{\psi} = \frac{S^2-\bar{t}^{\,2}}{S^2-(\bar{t}-1)^2}$$

Como en el caso anterior, el error medio cuadrático de este estimador es muy alto, además se presenta una alta cantidad de va lores inconscientes (lado derecho de la ecuación negativo) por tanto se descarta este estimador.

## 4.3. ECUACIONES INTEGRALES PARA OBTENER UN U.M.V.U.E.

Dado que la función $f_T(t|\psi)$ pertenece a una familia exponencial, $S = \sum t_i$ es un estadístico, suficiente mínimo completo luego, si $T$ es un estimador indeseado que es función de $s$, $T = g(S)$, entonces $T$ es U.M.V.U.E. encontrar tal estimador $T$, es equivalente a resolverla ecuación integral

$$\int_0^n g(s) f_s(s) ds = \psi$$

$$f_S(s) = \frac{1}{n}\left(\frac{\psi e^{\psi s/n}}{e^{\psi}-1}\right)^n H(s),$$

donde

$$H(s) = \sum_{k=0}^{n-1} \frac{n}{(n-1)!}\left[s^{n-1} - \binom{n}{1}(s-1)^{n-1} + \binom{n}{2}(s-2)^{n-1} - \ldots \right.$$

$$\left. \ldots + (-1)^k \binom{n}{k}(s-k)^{n-1}\right] I_{(k,k+1)}^{(s)},$$

esta ecuación se simplifica al considerar $g(s) = \frac{F(s)}{H(s)}$, $F(s)$ desconocida y la ecuación integral quedaría en la forma:

$$\int_0^n F(s) e^{\psi s} ds = \frac{n(e^{\psi}-1)^n}{\psi^{n-1}}$$

La última ecuación es una ecuación integral de Fredholm de primera especie. La forma general es:

$$\int_a^b K(x,y) f(y) dy = P(x)$$

$K \in L_2[a,b]\times[a,b]$, se denomina núcleo de la ecuación $P, f \in L_2[a,b]$.

El lado izquierdo de la ecuación es un operador lineal; desde el punto de vista de los operadores lineales la ecuación se escribiría en forma general como: $AX = Y$, si el operador inverso $A^{-1}$ existe la solución, sería $X = A^{-1}Y$; cuando el operador integral no es cerrado, es posible que no exista $A^{-1}$.

Las condiciones para la solubilidad de $AX = Y$ son:

i) $Y \in N(A^*)^\perp$

ii) $\Sigma\mu_n| <Y,h_n>|^2 < \infty$ donde $\mu_n$ son los valores propios, $h_n$ son los vectores propios ortonormales del operador $AA^*$.

Aquí $A^*$ es el operador transpuesto de $A$.

Cuando el núcleo es simétrico, $K(x,y) = K(y,x)$, el operador es simétrico, $A = A^*$, lo que ocurre con la ecuación, en estudio, por tanto los $\mu_n$ y $h_n$ de la condición ii) son vectores y valores propios de $A^2$ y la búsqueda de ellos para nuestro caso, es la búsqueda de los $\alpha$ tales que

$$F(\psi) = \alpha \int_1^n \frac{e^{n(s+\psi)} - e^{(s+\psi)}}{s+\psi} F(s)ds$$

esta búsqueda se intentó por dos métodos, por los determinantes de Fredholm y por el método de Rietz, pero los cálculos involucran términos complejos e inmanejables.

La alternativa es utilizar métodos de aproximación a la solución de nuestra ecuación original, el método utilizado requiere el concepto de operador inverso generalizado.

Sea $A:H_1 \to H_2$ un operador lineal acotado de un espacio de Hilbert $H_1$ en otro $H_2$, el operador inverso generalizado es aquel notado por $A^+$, con dominio $\mathcal{D}(A^+) = R(A) \oplus R(A)^{\perp}$; $A^+h_2 = h_1$ $h_1 \in H_1$, $h_1 \in S$, donde $S = \{h \in H_1 \mid Ah = Th_2\}$, $T$ es la proyección ortogonal de $H_2$ sobre $N(A)^{\perp}$ y tenemos que:

i) El operador integral tiene inverso generalizado.

ii) La ecuación $AX = Y$ tiene una solución aproximada de norma mínima, $X' = A^+Y$, $Y \in \mathcal{D}(A^+)$.

El método a seguir es el método de Cimmino presentado en Krasnov 1978 y Kolmogorov, Fomin 1978, Bazara, 1979.

$$X_1(t) = (I - \tfrac{2}{B} AA^*)X_o(t) + \tfrac{2}{B}(A^*Y)(t)$$

$$X_{n+1} = (I - \tfrac{2}{B} AA^*)X_n(t) + \tfrac{2}{B}(A^*Y)(t).$$

$\{X_n\}$ converge a la solución de norma mínima, tomando como aproximación inicial $X_o = 0$, $Y \in \mathcal{D}(A^+)$.

Como ocurrió en los anteriores casos, los cálculos para este caso son inmanejables.

Ante el fracaso de los métodos expuestos, se optó por resolver por métodos numéricos la ecuación de verosimilitud: $\bar{t} - E(\tau) = 0$ y comprobar el error medio cuadrático por medio de simulación, este resultó aceptable de modo que la estimación del parámetro $\psi$ se realizó por este medio.

## 5. Prueba de hipótesis Modelo I.

Las hipótesis a contrastar son:

$H_o$: $\psi$ es constante vs. $H_1$: $\psi$ es variable aleatoria para adecuar las hipótesis a la metodología de los test $C(\alpha)$ considere mos que $\psi = \psi_o + \sigma\sqrt{\xi}U$, $U$ variable aleatoria desconocida con $E(U) = 0$, $V(U) = 1$, entonces $\psi$ es variable aleatoria con $E(\psi) = \psi_o$, $V(\psi) = \xi\sigma^2$. Las hipótesis toman la forma:

$$H_o: \xi = 0 \quad \text{vs} \quad H_1: \xi > 0$$

de acuerdo con lo expuesto en los test $C(\alpha)$

$$r(t|\psi) = \frac{1}{f_2(t|\psi)} \cdot \frac{\partial^2 f_T(t|\psi)}{\partial\psi^2}, \quad Y(t|\psi) = \frac{LNf_T(t|\psi)}{\partial\psi}$$

La función de prueba $g$ vs: $g = r - aY$, donde

$$a = \frac{\text{Cov}(Y,r)}{V(Y)}$$

$$Z_n(\hat{\psi}) = \frac{\sum_{j=1}^{n}[Y(t_j|\psi_o) - aY(t_j|\hat{\psi}_o)]}{\sqrt{n}[V(g)]^{1/2}}$$

mediante cálculos rutinarios se muestra que

$$r(t|\psi) = t^2 - 2m_1 t + \frac{\psi}{e^{\psi-1}}(2m_1 - 1)$$

$$Y(t|\psi) = t - m_1, \qquad a = \frac{m_3 - 3m_1m_2 + 2m_1^2}{m_2 - m_1^2}$$

$$V(g) = V(r) - a\text{Cov}(r,Y).$$

Este test se implementó en la historia de accidentes de los corteros de caña del Ingenio Risaralda, tomando como intervalo unitario de observación el año 1984; se realizaron tres pruebas para tamaños de muestra 10, 20, 30, 40, 50, 60, 70 y en todas no se rechaza $H_0$ lo que indica no aleatoriedad en el contagio por accidentes.

## 6. Estimación de Parámetros Modelo II.

La función de densidad base de la estimación en este módelo es:

$$f_T(t|\psi,\nu) = \frac{\psi}{(1+\nu)^{\psi/\nu}-1}(1+\nu t)^{(\psi/\nu)-1}$$

La función de verosimilitud esta dada por:

$$L(\psi,\nu) = \left(\frac{\psi}{(1+\nu)^{\psi/\nu}-1}\right)^n \prod_{k=1}^{n}(1+\nu t_k)^{(\psi/\nu)-1}$$

y después de reducciones se obtiene la ecuación

$$(1) \qquad \frac{n}{\nu} - \frac{n\psi(1+\nu)^{\psi/\nu-1}}{\nu[(1+\nu)^{\psi/\nu}-1]} + (\psi/\nu-1)\sum_{k=1}^{n}\frac{t_k}{1+\nu t_k} = 0\ .$$

La cual no se puede resolver algebraicamente; para aproximar la solución de esta ecuación se fijo la variable $\nu$ y se resuelve la ecuación para $\psi$, tratando $\nu$ como constante, la ecuación queda en la forma:

$$\frac{n}{\nu} - \frac{n\psi\, a^{\psi}}{(1+\nu)\nu(a^{\psi}-1)} + (\psi/\nu-1)\sum_{k=1}^{n}\frac{t_k}{1+\nu t_k} = 0; \quad a = (1+\nu)^{1/\nu},$$

aproximando por polinomios de Taylor $a^{\psi}$ y resolviendo se obtiene una solución de $\psi$ que depende de $\nu$ y los $t_i$; para establecer la solución, se simulan muestras $t_1, t_2, \ldots, t_n$ con valo res de $\psi$, $\nu$ prefijados en la simulación, luego se toman valores de $\nu$ cercanos al valor utilizado en la simulación, y así se calcula $\psi$ de la ecuación anterior.

Los resultados obtenidos mostraron una posible relación lineal entre los dos parámetros $\psi$ y $\nu$, dado que estos resultados se obtuvieron utilizando aproximaciones, se realizó un procedimiento similar con la ecuación original, los resultados mostraron, efectivamente, relación lineal entre los parámetros: $\psi = \beta_o + \beta, \nu$, pero aún algo más, cuando $0 < \nu < 1$ el modelo es $\psi = -0.005 + \nu$; no importa que tamaño de muestra se tome, o que valores de $\psi$ y $\nu$ se utilicen en la simulación; así que en este caso, la estimación se reduce a la de $\nu$, la ecuación toma la forma:

$$(2) \qquad n - \frac{n(\nu - 0.005)}{(1+\nu)(1+\nu)^{0.005/\nu}} - 0.005 \sum_{k=1}^{n} \frac{t_k}{1+\nu t_k} = 0$$

aún asi no es posible despejar $\nu$, pero el término $(1+\nu)^{0.005/\nu} \simeq 1$ y al sustituir y reducir se llega a $\sum_{k=1}^{n} \frac{1}{1+\nu t_k} = 0$ lo cual no es posible pues tanto $\nu$ como los $t_k$ son positivos; se procedió entonces por dos lados: aproximar el lado izquierdo de la ecuación (2) polinomios de Taylor y resolver; el otro aproximar el término $(1+\nu)^{0.005/\nu}$ por polinomios de Taylor y aplicar el método de los momentos; en los dos casos, al simular muestras, $t_1, t_2, \ldots, t_n$, solo aproximadamente el 50% de las veces se obtuvieron estimadores aceptables.

En el caso más general, $\nu > 1$, la relación lineal no muestra ninguna uniformidad lo que impide su utilización.

El método utilizado fue, de nuevo, optimizar la función de verosimilitud por métodos numéricos, en esta ocasión el método denominado de coordenadas cíclicas fue el que se aplicó.

## 7. Test de hipótesis Modelo II

Las hipótesis a contrastar son las mismas que en el Modelo I y las consideraciones sobre adecuación de hipótesis para la metodología $C(\alpha)$, las mismas que en el modelo I; aquí el parámetro $\theta = (\psi,\nu)$ y la función de prueba está dada por:
$g = r - bY_1 - cY_2$ donde:

$$r = \frac{1}{f_T(t|\psi,\nu)} \frac{\partial^2 f_T(t|\psi,\nu)}{\partial\psi^2} \; ; \quad Y_1 = \frac{\partial LN f_T(t|\psi,\nu)}{\partial\psi}$$

$$Y_2 = \frac{\partial LN f_T(t|\psi,\nu)}{\partial\nu}$$

$$b = \frac{V(Y_2)\operatorname{Cov}(r,Y_1) - \operatorname{Cov}(Y_1,Y_2)\operatorname{Cov}(Y_2,r)}{V(Y_1)V(Y_2) - [\operatorname{Cov}(Y_1,Y_2)]^2}$$

$$c = \frac{\operatorname{Cov}(Y_2,\nu) - b\operatorname{Cov}(Y_1,Y_2)}{V(Y_2)}$$

$$Z_n(\hat\theta) = \frac{\sum_{i=1}^{n} g(t_i,\hat\theta)}{\sqrt{nV(g)}}$$

$$g(t_i,\hat\theta) = r(t_i,\hat\theta) - bY_1(t_i,\hat\theta) - cY_2(t_i,\hat\theta).$$

El test se aplicó a la misma población del Modelo I, para tamaños de muestra 30, 40, 50, 60, 70, 80, 90, a un nivel de 5% y 1% rechazándose $H_o$ en todos los casos (menos uno), lo cual indica aleatoriedad en el contagio, en contraste con el Modelo I, parece ser que el efecto del tiempo introduce aleatoriedad en el contagio.

*

## BIBLIOGRAFIA


Bates, G.E., (1955), *Joint distributions of time intervals for the ocurrence of sucessice accidents in a generalized Polya Scheme.* Ann. Math. Statistics (26), pp.705-770.

Bates, G.E. and Neyman, J., (1952), *Contribution to the theory of accident proneness. True o false contagion.* Univ. of Glif. Public. in Stadistics (1), pp.255-276.

Bazora, M.S. and Shetty, C.M., (1979), *Nonlinear programing, Theory and Algorithms.* John Wiley and Sons.

Cramer, Harald, (1968), *Métodos Matemáticos de Estadística.* Aguilar S.A. de Ediciones. Cuarta Edición.

Feller, William, (1968), *An Introduction to probability theory and its applications.* Wiley Internacional Edition. Third Edition.

Graybill, Franklim, A., (1978), *Theory and Applications of the Linear Models.* Duxbury Press, Massachusetts, U.S.A.

Greenwood, M. and Yolr, G.V., (1920), *An inquiry into the nature of frequency distributions, representative of multiple happenings with particular reference to the ocurrence of multiple attachs of disease as repeated accidents.* J. R. Statist. Soc. (83) pp.255-279.

Groetsch, C.W., (1977), *Generalized, inverses of linear operators representation and arpoximations.* Marcedl Dekker Inc.

Kammerer, W.J., (1972), *Iterative methods for best approximate solutions of linear integral equations of the first and second kind.* J. Math. Anal. Appl. (40), pp.547-573.

Kendall, H.G., Stuart, (1967), *The advances theory of Statistics.* Ch. Griffin & Company Limited. London.

Kolmogorov, M. Foming, S., (1978), *Introdución a la Teoría de funciones y al análisis funcional.* Editorial MIR, Moscú 3a. Ed.

Krasnov, M., Kiseliov, A., (1978), *Makarenko, G. Ecuaciones Integrales.* Editorial MIR. Moscú, 3a. Ed.

Lehmann, E.L., (1959), *Testing Statistical Hypotheses.* Wiley Publications inf Mathematical Statistics.

Lewis, A.W., Peter, (1972), *Stochastic Point Processes: Statistical analysis, theory and applications.* Wiley Series in Probabilities and Mathematical Statistics.

Mood M., Alexander; Grayvill, A., Franklim and Boes, C.D., *Introduction to the theory of Statistics.* McGraw-Hill. Series in Probability and Stadistics. 3a. Ed.

Neyman, J., (1959), *Optimal asymtotic test of composite statistical hypotheses.* The Harald Cramer Volume and Wiksell, Stocolm and Wiley and sons, New York. pp.213-234.

Neyman, J. and Scott, E., (1972), *On the use of optimal C(α) test of composite hypotheses.* Bull de L'institute international de Statistique (41), pp.477-496.

Petrowski, I.G., (1977), *Lecciones de la teoría de Ecuaciones integrales.* Editorial MIR Moscú.

Polya, G., (1930), *Sur quelques points de la theorie des probailities.* Ann. Inst. Henri Poincare (1), pp.117-161.

Sobol, I.M., (1976), *Método de Montecarlo.* Edit.MIR. Moscú.

Sánchez, Velasco, (1991), *Curso Básico de Algebra Lineal.* 6a. Ed. Editorial Comex, Bogotá.