MFN-008
TM-008

Pontificia Universidad Católica de Chile
Facultad de Matemática – Depto. de Estadística
Programa de Magíster en Estadística

MODELOS DE SERIES CRONOLOGICAS PARA FAMILIA EXPONENCIAL Y APLICACIONES

Universidad Mayor de San Andres
Facultad de Ciencias Puras y Naturales
CARRERA DE MATEMATICA
BIBLIOTECA
LA PAZ - BOLIVIA

DIVISION BIENES E INVENTARIOS
VERIFICADO
Fecha 2-2-96
H. M. R
M S

Trabajo de Tesis de Magister Realizado Por

Porfirio Suñagua S.

Bajo la Orientación del

Prof. Pablo Marshall R.

Santiago, 21 de Diciembre de 1995

A Gladys, Marcelo y Rodrigo

También a: Benigno, †Policarpio y †Cipriana

AGRADECIMIENTOS

Este trabajo de tesis, como producto de haber terminado mis estudios en el Programa de Magíster en Estadística en la Facultad de Matemáticas de la Pontificia Universidad Católica de Chile, fue posible por la colaboración de distintas instituciones y personas a los cuales les quiero agradecer.

En primer lugar a mi familia en La Paz–Bolivia, quienes con mucha comprensión me apoyaron y me ayudaron en todo momento para que pueda realizar mis estudios en Chile. A mi tutor de tesis, Prof. Pablo Marshall, de quién he recibido muchos conocimientos particularmente en el área de series temporales, y sin su participación no hubiera sido posible la conclusión satisfactoria del presente trabajo de tesis. Al Banco Interamericano de Desarrollo (BID) con fondos del gobierno de Japón, por haberme otorgado una beca de estudios durante 1 año y medio, con el cuál fué superada el aspecto financiero de mi estadía en Chile y los pagos correspondientes a la Universidad durante ese período. Al Dirección de Post–Grado de la Pontificia Universidad Católica de Chile, por haberme brindado todo el apoyo necesario durante los dos años de mis estudios y muy particularmente el último semestre en donde me beneficiaron con una ayuda económica destinada a la elaboración de tesis. A la Carrera de Matemáticas de la Universidad Mayor de San Andrés de La Paz–Bolivia, por haberme otorgado una licencia por todo el periodo de tiempo que duraron mis estudios de Post–Grado en Chile y por haberme brindado todo el apoyo necesario y muy especialmente la Jefatura de Carrera y al Honorable Consejo de Carrera. A mis profesores de los distintos ramos en el Departamento de Estadística, de quienes he recibido en sus enseñanzas muchos conocimientos con los cuales mi desenvolvimiento como profesional en el futuro será fructífera. Al Departamento de Estadística, por haberme brindado una oportunidad de realizar mis estudios de Post–Grado y por haberme dado toda la atención necesaria en todo momento, especialmente en los aspectos administrativos. A los Directores de Post–Grado, primero Prof. Guido del Pino y luego Prof. Guillermo Marshall, de quienes he recibido mucho apoyo moral y material y que ellos han hecho posible de que mi estadía en la Universidad sea muy provechosa. A mis compañeros de mi promoción, con quienes hemos formado un grupo académico y social muy bueno y que han hecho posible que estuviera como en mi país y también a los administradores de la sala de computación de la Facultad de Matemáticas. Al Club Deportivo Universidad Católica, Gerencia de Marketing, por haberme proporcionado los datos para una de las aplicaciones. A la Corporación de Televisión Universidad Católica de Chile, Canal 13 de Santiago, por haberme proporcionado los datos de los raiting de televisión de los programas de Noticieros. Finalmente de manera especial a los profesores Guido del Pino y Pilar Iglesias del Departamento de Estadística, por sus sugerencias y comentarios antes y durante la presentación de este trabajo como parte de la comisión examinadora.

21 de Diciembre de 1995
Porfirio Suñagua S.

Contenidos

Lista de Tablas v

Lista de Figuras vi

Resumen vii

1 FAMILIA EXPONENCIAL Y ANALISIS CONJUGADO 1
- 1.1 Familia Exponencial 1
- 1.2 Familia Conjugada 3
  - 1.2.1 A Priori Conjugada 3
  - 1.2.2 Análisis Conjugado 5
  - 1.2.3 Ejemplos 7

2 MODELOS DE SERIES DE TIEMPO PARA FAMILIA EXPONENCIAL 13
- 2.1 Preliminares 13
- 2.2 Especificación del Modelo 14
- 2.3 Estimación 17
- 2.4 Predicción 20
- 2.5 Ejemplos 22
- 2.6 Suavizamiento 26
- 2.7 Función de Verosimilitud 27
- 2.8 Residuo de Pearson 27

3 APLICACION I: GOLES MARCADOS POR UN EQUIPO DE FUTBOL 28
- 3.1 Análisis del Problema 28
- 3.2 Un Modelo Dinámico Poisson–Gama 29
  - 3.2.1 El Número de Goles de la U. Católica 29
  - 3.2.2 Estimación 32
  - 3.2.3 Nivel de la Tendencia en Línea y Suavizada 35
  - 3.2.4 Predicción 37
  - 3.2.5 Residuo de Pearson 37
  - 3.2.6 Función de Verosimilitud 39
- 3.3 Conclusiones 40

4 APLICACION II: RATING DE LA TELEVISION (TVR) 42
4.1 Análisis del Problema 42
4.2 Un Modelo Dinámico Binomial–Beta 44
4.2.1 TVR de los Noticieros 44
4.2.2 Estimación 45
4.2.3 Nivel de la Tendencia en Línea y Suavizada 49
4.2.4 Predicción 53
4.2.5 Residuo de Pearson 55
4.2.6 Función de Verosimilitud 55
4.3 Conclusiones 57

Apéndice 59

A CALCULOS AUXILIARES 59
A.1 Momentos en el Modelo Poisson–Gama 59
A.2 Momentos en el Modelo Beta–Binomial 61
A.3 Momentos en el Modelo Binomial Negativa–Beta 62

B SUBRUTINAS COMPUTACIONALES EN S–PLUS 63
B.1 Rutinas Generales 63
B.2 Rutinas Particulares Para el Modelo Poisson–Gama 74
B.3 Rutinas Particulares Para el Modelo Binomial–Beta 78

Bibliografía 83

Lista de Tablas

1.1 Distribuciones a Priori Conjugadas 5

3.1 Frecuencias Relativas y Probabilidades 29
3.2 Estimación de la Tendencia 35
3.3 Probabilidades del Próximo Partido 37
3.4 Devianzas 40
3.5 Goles de los Clásicos Universitarios 41

4.1 Estimación de la Tendencia en Línea y Suavizada 49
4.2 Predicciones de los Ratings 54
4.3 Varianzas Estimadas 55

Lista de Figuras

3.1 Probabilidades Poisson 30
3.2 Número de Goles de U. Católica 31
3.3 Promedio de Goles y Tendencia (en línea) 36
3.4 Promedio de Goles y Tendencia (suavizada) 36
3.5 Probabilidad del Próximo Partido 38
3.6 Residuos de Pearson 39

4.1 Rating de Programas de Noticias 43
4.2 Estimación en Línea: Programa TELETRECE 50
4.3 Estimación Suavizada: Programa TELETRECE 50
4.4 Estimación en Línea: Programa 24 HORAS 51
4.5 Estimación Suavizada: Programa 24 HORAS 51
4.6 Estimación en Línea: Programa MEGANOTICIAS 52
4.7 Estimación Suavizada: Programa MEGANOTICIAS 52
4.8 Residuos de Pearson: Programa TELETRECE 56
4.9 Residuos de Pearson: Programa 24 HORAS 56
4.10 Residuos de Pearson: Programa MEGANOTICIAS 56

Resumen

DIVISION BIENES E INVENTARIOS
VERIFICADO
Fecha
H. M. R.
M. S. A.

Observaciones en la forma de series cronológicas o series temporales son comunes en Estadística. Una parte significativa de la literatura especializada se concentra en el análisis de este tipo de datos. Es así como la estimación de tendencias, efectos estacionales y efectos cíclicos, conjuntamente con la predicción de observaciones futuras, constituyen los objetivos fundamentales en estos estudios.

Las distintas clases de modelos que se han utilizado tradicionalmente para el análisis y la predicción de series cronológicas tienen en común que suponen que la distribución de probabilidades asociada a las observaciones es Normal, o aproximadamente Normal. Los modelos ARIMA, desarrollados por Box y Jenkins(1976), y los todavía nuevos modelos estructurales desarrollados por Harvey(1989) en base al filtro de Kalman suponen distribución Normal para las observaciones o restringen el análisis a los dos primeros momentos, y por tanto a estimaciones y predicciones lineales.

A pesar de que el mayor interés en el desarrollo de nuevas metodologías ha estado enfocado a estos métodos, muchas series cronológicas que se encuentran en el mundo real tienen observaciones provenientes de otras distribuciones y no existe una metodología universalmente aceptada que pueda ser utilizada en estos casos.

En este trabajo se presenta una metodología para el análisis de series cronológicas con observaciones provenientes de una distribución cualquiera en la familia exponencial. Para esto se utiliza el desarrollo en la escuela clásica de Harvey y Fernandes(1989), que también aparece en Harvey(1989), y los desarrollados en la escuela bayesiana de West y Harrison(1989), West(1985a), West, Harrison y Migon(1985), y de Pole, West y Harrison(1994). El desarrollo que se sigue aquí es más cercano al de la escuela bayesiana. En algún sentido, la extensión de los modelos estructurales que se basan en el filtro de Kalman a modelos para toda las distribuciones de la familia exponencial se asemeja a la extensión desarrollada principalmente por Nelder y Wedderburn(1972), Baker y Nelder(1983), y por McCullagh y Nelder(1989) de los modelos lineales a los modelos lineales generalizados.

Las contribuciones de este trabajo son: La presentación de una metodología completa para modelar series de tiempo univariadas cuyas observaciones tienen distribución cualquiera en la familia exponencial en el sentido de complementar con aportes adicionales a las descritas en West y Harrison(1989) con las ideas de estimación de hiperparámetros, la estimación suavizada y defi-

nición de residuos. La implementación de las rutinas computacionales en el lenguaje S–Plus que permiten ajustar este tipo de modelos. Otro aporte importante del trabajo es el modelamiento de los goles marcados por un equipo de fútbol contra otro equipo en particular en mismo tipo de torneos, los cuales se pueden considerar como procesos provenientes de la distribución Poisson. Otra aplicación imporante es el análisis de datos correspondientes a los ratings de programas de televisión de la Ciudad de Santiago, los cuales son considerados como procesos provenientes de la distribución Binomial. Finalmente se presentan algunos resultados relacionados con cálculos de momentos que no son comunmente conocidos.

El trabajo, aparte del presente resumen, se organiza de la siguiente manera. En el primer capítulo se presentan resultados generales en relación a la familia exponencial de distribuciones y a la familia conjugada de distribuciones. En esto se sigue muy de cerca el planteamiento en Bernardo y Smith(1994). En algún sentido, esto constituye el sustento teórico que se utiliza luego en los aspectos más metodológicos.

En el Capítulo 2 se presenta la metodología general para modelar series cronológicas con observaciones en la familia exponencial. En esta parte del trabajo se sigue muy de cerca los resultados de West y Harrison(1989), West, Harrison y Migon(1985), West(1985a) aunque una cantidad considerable de resultados han sido derivados.

Los Capítulos 3 y 4 están dedicados a dos aplicaciones en las cuales se muestran las bondades de la metodología propuesta. En la primera de estas aplicaciones, en el Capítulo 3, se estudia la evolución dinámica del número de goles marcados por el equipo de fútbol de la Universidad Católica en partidos del campeonato nacional de Chile jugados contra el equipo de la Universidad de Chile entre 1939 y 1994. La naturaleza de este tipo de observaciones sugiere que pueden ser considerados como una realización de una variable aleatoria con distribución Poisson. Por otra parte, la natural dependencia de las observaciones a través del tiempo aconseja el uso de una clase de modelos de series de tiempo para estudiar y predecir el comportamiento futuro del número de goles marcados. El número de goles marcados en años consecutivos están correlacionados y no pueden ser tratados como realizaciones independientes.

En la segunda aplicación, correspondiente al Capítulo 4, se modelan ratings de programas de televisión a partir de las observaciones del sistema *people meter* que se utiliza en Chile para medir los ratings de los programas de televisión. El modelo propuesto en este caso toma en cuenta que las observaciones provienen de una distribución Binomial y utiliza las correlaciones de observaciones vecinas para predecir en el corto plazo el comportamiento de los ratings.

Un aporte adicional que se realiza en este trabajo consiste en un conjunto de rutinas computacionales para estimar los distintos modelos y tener predicciones. Estas rutinas están escritas en el lenguaje S–plus, y se encuentran en el Apéndice B.

Un análisis más sofiticado se puede estudiar incorporando al modelo otros factores adicionales que no estan consideradas en el modelo propuesto, como variables dummy, variables de

intervención, variables exógenas. En el caso de los goles, la dependencia de la pericia del otro equipo, así como también un estudio multivariado con el fin de capturar las correlaciones de los goles marcados por el otro equipo o el rendimiento de todos los equipos que compiten en el mismo torneo. Para el caso de los rating de la televisión también se podría considerar un modelo multivariado Multinomial para incorporar al modelo las correlaciones existentes entre los ratings de distintos programas que se están transmitiendo a la misma hora.

Capítulo 1

FAMILIA EXPONENCIAL Y ANALISIS CONJUGADO

En este capítulo se define la familia exponencial de distribuciones. En primer lugar se define la familia uniparamétrica y k–paramétrica, la representación canónica de estas familias. Luego se define la familia conjugada, y se describe una metodología del análisis conjugado siguiendo de cerca los procedimientos desarrollados por Bernardo y Smith(1994, Cap.4). Al final del capítulo se presenta algunos ejemplos que se usarán posteriormente.

1.1 Familia Exponencial

Definición 1.1.1 *Una densidad o función de probabilidad (fdp) $p(y|\theta)$, donde $\theta \in \Theta \subset \mathbb{R}$ de una variable aleatoria (v.a.) definida sobre $\mathcal{Y}$ se dice que pertenece a la familia exponencial (FE) uniparamétrica si*

$$p(y|\theta) = f(y)g(\theta)\exp\{c\xi(\theta)h(y)\} \tag{1.1}$$

para algunas funciones conocidas $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ y $\xi(\cdot)$ y una constante c tal que

$$[g(\theta)]^{-1} = \int_{\mathcal{Y}} f(y)\exp\{c\xi(\theta)h(y)\}dy < \infty \tag{1.2}$$

la familia se llamará *regular* si $\mathcal{Y}$ no depende de θ. En caso contrario se llamará *no–regular*. Algunas de las variables aleatorias que pertenecen a esta familia son: *Binomial, Poisson, Exponencial, Gama, Normal, Binomial–Negativa, Uniforme* y *Exponencial–Truncada.* Los dos últimos casos mencionados corresponden a la clase de familia exponencial no–regular.

Por otra parte, la aplicación $\eta = c\xi(\theta)$ se conoce como la función de *enlace natural*, y en ese caso η es llamado *parámetro natural*, de hecho, podemos tomar otras funciones de enlace, las cuales pueden ser apropiadamente especificadas por algún criterio de buen ajsute a los datos por el modelo en consideración.

Definición 1.1.2 *Una densidad o función de probabilidad $p(y|\theta)$, $y \in \mathcal{Y}$, $\theta \in \Theta \subset \mathbb{R}^k$, se dice que pertenece a la Familia Exponencial k-paramétrica si*

$$p(y|\theta) = f(y)g(\theta)\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)h_i(y)\right\} \tag{1.3}$$

para algunas funciones conocidas $f(\cdot)$, $g(\cdot)$, $h = (h_1, \cdots, h_k)$ y $\xi(\theta) = (\xi_1(\theta), \cdots, \xi_k(\theta))$ y una constante c tal que

$$[g(\theta)]^{-1} = \int_{\mathcal{Y}} f(y)\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)h_i(y)\right\} dy < \infty \tag{1.4}$$

como en el caso uniparamétrico, la familia se llamará *regular* si $\mathcal{Y}$ no está en dependencia del espacio paramétrico. En otro caso se llamará *no-regular.*

Las variables aleatorias, como *Normal, Multinomial, Uniforme*, etc. ilustran esta definición para ambos casos de regularidad. De hecho, las variables aleatorias con distribución en la familia exponencial ya sea uniparámetrica o multiparamétrica que tienen mayor aplicación, son aquellas de la clase regular.

Definición 1.1.3 (Representación Canónica de la FE) *La densidad o la función de probabilidad $p(y|\psi)$ de la familia exponencial se dice que es canónica si*

$$p(y|\psi) = p(y|a, b, \psi) = b(y)\exp\{y'\psi - a(\psi)\}, \quad y \in \mathcal{Y} \tag{1.5}$$

Donde y' es la transpuesta de y. Esta definición está derivada a partir de la Definición 1.1.2, vía transformaciones.

$$y = (y_1, \cdots, y_k)^T \qquad \psi = (\psi_1, \cdots, \psi_k)$$

$$y_i = h_i(y) \qquad \psi_i = c_i\xi_i(\theta) \qquad i = 1, 2, \cdots, k.$$

Generalmente la forma que adoptaremos en los próximos capítulos para el caso univariado de esta representación será

$$p(y|\eta, V) = b(y, V)\exp\left\{\frac{y\eta - a(\eta)}{V}\right\} \tag{1.6}$$

donde η es llamado parámetro natural o de localización, V se llama parámetro de escala, el cual habitualmente se toma positivo y para algunas distribuciones éste resulta conocido. A veces es conveniente trabajar con la reparametrización $\phi = \frac{1}{V}$, en este caso, ϕ es llamado parámetro de precisión.

Más aún, casi todas las distribuciones comunmente conocidas en Estadística se pueden escribir en esta forma, lo que facilita los cáculos posteriores en su aplicación. En la Ecuación (1.6) para algunas variables aleatorias, en lugar de y se podría tener una función $\boldsymbol{Y}(y)$, como resulta en el caso de la variable aleatoria Binomial.

DIVISIÓN BIENES E INVENTARIOS
VERIFICADO
Fecha.....
H. M. R.
M. S.

1.2 Familia Conjugada

1.2.1 A Priori Conjugada

Definición 1.2.1 *La densidad a priori conjugada para $\theta \in \Theta$, con respecto a una verosimilitud $p(y|\theta)$ con estadístico suficiente $t = t(y) = \{n, S(y)\}$ k-dimencional, es*

$$p(\theta|\tau) = \frac{p(S = (\tau_1, \cdots, \tau_k)|\theta, n = \tau_0)}{\int_\Theta p(S = (\tau_1, \cdots, \tau_k)|\theta, n = \tau_0)d\theta}, \qquad \tau \in \mathcal{T} \tag{1.7}$$

donde $\tau = (\tau_0, \tau_1, \cdots, \tau_k)$ y

$$\mathcal{T} = \left\{\tau : \int_\Theta p(S = (\tau_1, \cdots, \tau_k)|\theta, n = \tau_0)d\theta < \infty\right\}.$$

La familia $\{p(\theta|\tau)\}_{\tau\in\mathcal{T}}$ tal que la operación bayesiana sea cerrada en el sentido que la posteriori de $p(y|\theta)$ y $p(\theta|\tau)$ pertenece a la misma familia, esto se conoce como la familia a priori conjugada natural asociada a la verosimilitud $p(y|\theta)$ y n es el tamaño muestral con el cual se tiene el estadístico $S(y)$.

Un resultado muy importante, cuyos detalles de fundamentación se puede encontrar en Bernardo y Smith(1994 Cap.4) que textualmente dice que, tomando la definición anterior, la verosimilitud para el cual la familia conjugada a priori existe, son aquellos que corresponden a la familia exponencial paramétrica generalizada. Así, dados $f(\cdot)$, $g(\cdot)$, $h(\cdot)$, $\xi(\cdot)$ y una constante c como en la Definición 1.1.2 se tiene

$$p(y|\theta) = f(y)g(\theta)\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)h_i(y)\right\}, \qquad y \in \mathcal{Y} \tag{1.8}$$

$$[g(\theta)]^{-1} = \int_{\mathcal{Y}} f(y)\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)h_i(y)\right\}dy < \infty. \tag{1.9}$$

Proposición 1.2.1 (Familia Conjugada para FE Regular) *Sea $Y = (Y_1, \cdots, Y_n)$ una muestra de una variable aleatoria (v.a.) con distribución en la familia exponencial regular con densidad o función de probabilidad*

$$p(\boldsymbol{y}|\theta) = \prod_{i=1}^{n} f(y_i)\ [g(\theta)]^n \exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)\sum_{j=1}^{n} h_i(y_j)\right\} \tag{1.10}$$

entonces la densidad de la familia a priori conjugada para θ tiene la forma

$$p(\theta|\tau) = [K(\tau)]^{-1}[g(\theta)]^{\tau_0}\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)\tau_i\right\}, \qquad \theta \in \Theta \tag{1.11}$$

donde $\tau = (\tau_0, \tau_1, \cdots, \tau_k)$ es tal que

$$K(\tau) = \int_\Theta [g(\theta)]^{\tau_0}\exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)\tau_i\right\}d\theta \ < \ \infty \tag{1.12}$$

llamada la constante normalizadora.

Prueba. Por el criterio de factorización de Neyman, el estadístico suficiente para $\xi(\theta)$ tiene la forma siguiente

$$t_n(Y_1,\cdots,Y_n) = \left[n, \sum_{j=1}^{n} h_1(Y_j), \cdots, \sum_{j=1}^{n} h_k(Y_j)\right] = [n, S(\boldsymbol{Y})] \tag{1.13}$$

Así, para cualquier τ tal que $\int_\Theta p(\theta|\tau)d\theta < \infty$ una densidad a priori conjugada, por la Definición 1.2.1 tiene la forma

$$\begin{aligned} p(\theta|\tau) &\propto p(S_1(\boldsymbol{y}) = \tau_1, \cdots, S_k(\boldsymbol{y}) = \tau_k|\theta, n = \tau_0) \\ &\propto [g(\theta)]^{\tau_0} \exp\left\{\sum_{i=1}^{k} c_i\xi_i(\theta)\tau_i\right\}. \end{aligned}$$

Luego, dividiendo por la costante normalizadora, se obtiene el resultado deseado.

Aquí es muy importante notar que, bajo condiciones de regularidad de la verosimilitud, aunque ésta no sea una de las comunmente conocidas, la Proposición 1.2.1 nos permite encontrar analíticamente la a priori conjugada. Aunque las operaciones a veces podrían resultar complicados, no solamente en la obtención de la distribución a priori conjugada si no también en los cálculos posteriores. Por ejemplo en el cálculo de los momentos.

Para el caso de la verosimilitud en la FE canónica regular, presentamos en las siguientes líneas, la aplicación del procedimiento descrito anteriormente.

Sean $Y_1, Y_2, \cdots, Y_n$ *iid* $p(y|\eta, V)$, $(V > 0$ conocido$)$ $\phi = \frac{1}{V}$

$$p(y|\eta,\phi) = b(y,\phi)\exp\{\phi[y\eta - a(\eta)]\} \tag{1.14}$$

$$\begin{aligned} p(y_1, y_2, \cdots, y_n|\theta) &= \prod_{i=1}^{n} b(y_i,\phi)\exp\{\phi[y_i\eta - a(\eta)]\} \\ &= \prod_{i=1}^{n} b(y_i,\phi)\; e^{-\phi a(\eta)n}\exp\left\{\eta\phi\sum_{i=1}^{n} y_i\right\} \end{aligned}$$

Por la Proposición 1.2.1, la densidad a priori conjugada está dado por

$$\begin{aligned} p(\theta|\tau_0,\tau_1) &\propto e^{-\phi a(\eta)\tau_0}\exp\{\eta\phi\tau_1\} \\ &\propto \exp\{\phi[\eta\tau_1 - a(\eta)\tau_0]\}. \end{aligned}$$

Si asumimos la existencia de la constante normalizadora $K(\tau_0,\tau_1)$ para un rango apropiado para los parámetros, se tiene que

$$p(\eta|r,s) \propto \exp\{r\eta - sa(\eta)\}$$

donde $r = \phi\tau_1$, $s = \phi\tau_0$. Luego la a priori conjugada para η, asociado a la verosimilitud dada en la Ecuación (1.14), está dado por

$$p(\eta|\tau_0,\tau_1) = p(\eta|r,s) = c(r,s)\exp\{r\eta - sa(\eta)\}, \qquad s > 0 \tag{1.15}$$

donde c(r,s) es el inverso de la constante normalizadora.

Así esta última aplicación ilustra la existencia de la a priori conjugada para las variables aleatorias habitualmente de interés. Del mismo modo se puede aplicar la misma metodología para encontrar la a priori conjugada para aquellas verosimilitudes que no tienen un nombre especial. En algunos casos, aunque la verosimilitud es conocida, su a priori conjugada no siempre resultará uno conocido, tal es el caso de la verosimilitud *Beta*.

Una extensión natural, se puede tener para dimenciones mayores del parámetro, como el caso de la *Normal Multivariada*, *Multinomial* entre otros. De todas maneras, en la Tabla 1.1 presentamos un listado de las a priori conjugadas asociados a sus correspondientes verosimilitudes, los mismos que son frecuentemente conocidos en la literatura estadística. Mayores detalles se puede encontrar en De Groot (1970 Cap.9) y Bernardo y Smith (1994,Cap.5).

Tabla 1.1: **Distribuciones a Priori Conjugadas: p, r, m y V conocidos y $\boldsymbol{R}$, τ matrices de precisión**

Verosimilitud	A Priori Conjugada	Espacio Paramétrico
$Ber(\theta)$	$Beta(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$Poisson(\theta)$	$Gama(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$N(\mu,\lambda)$	$N(\mu\|\gamma,\lambda(2\alpha-1))Gama(\lambda\|\alpha,\beta)$	$\alpha>\frac{1}{2},\ \ \beta>0,\ \ \gamma\in\mathbb{R}$
$Gama(p,\theta)$	$Gama(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$Exp(\theta)$	$Gama(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$BinNeg(\theta,r)$	$Beta(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$Bin(n,\theta)$	$Beta(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$U[0,\theta]$	$Pareto(\theta\|\alpha,\theta_0)$	$\alpha>0,\ \ \theta_0>0$
$Geo(\theta)$	$Beta(\theta\|\alpha,\beta)$	$\alpha>0,\ \ \beta>0$
$b(y,V)\exp\{V^{-1}[y\eta-a(\eta)]\}$	$c(r,s)\exp\{r\eta-sa(\eta)\}$	$s>0$
$U[\theta_1,\theta_2]$	$ParetoBil(\theta_1,\theta_2\|r_1,r_2,\alpha)$	$r_1<r_2,\ \ \alpha>0$
$Mult(n,\pi_1,\cdots,\pi_k)$	$Dirichlet(\pi_1,\cdots,\pi_k\|\alpha_1,\cdots,\alpha_k)$	$\forall i,\alpha_i>0$
$N(\vec{\boldsymbol{\mu}},\boldsymbol{R})$	$\vec{\boldsymbol{\mu}}\|\boldsymbol{R}=r\sim N(\mu,\nu r)$	$\nu>0,\ \ \mu\in\mathbb{R}^k$
	$R\sim Wishart(\alpha,\tau)$	g.l. $\alpha>k-1$

1.2.2 Análisis Conjugado

Aunque la definición de la distribución a priori conjugada es más general que para variables aleatorias con distribución en la familia exponencial, en esta aplicación sin perder generalidad

nos limitaremos solamente a hacer el análisis conjugado para variables cuya distribución de probabilidades se escribe en la forma adoptada en la Ecuación (1.6). Es decir, si y es la realización de una variable aleatoria Y cuya distribución de probabilidades $p(y)$ pertenece a la familia exponencial canónica con parámetros η y ϕ, con $\phi > 0$. Además suponemos que el parámetro ϕ es conocido.

Ahora, supongamos que η también es considerado como una variable aleatoria con distribución de probabilidades $p(\eta)$, de modo que la distribución de probabilidades de Y en realidad es condicional en un valor del parámetro η y será mejor representarla por $p(y|\eta)$.

Por el teorema de Bayes se puede obtener la distribución de probabilidades de η condicional en la observación $Y = y$, que es la a posteriori de η.

$$p(\eta|y) \propto p(y|\eta)p(\eta). \tag{1.16}$$

Por otra parte, la distribución marginal de Y llamado también la distribución predictiva se puede obtener vía

$$p(y) = \int p(y|\eta)p(\eta)d\eta, \tag{1.17}$$

donde la integración se hace sobre todo el espacio paramétrico para η. De hecho los cálculos en (1.16) y (1.17) pueden resultar muy complicados dependiendo de la naturaleza de la distribución a priori. En el caso de normalidad la resolución es bastante sencilla, mientras que, aún es resoluble analíticamente cuando la distribución de las observaciones pertenecen a la familia exponencial siempre que su a priori pertenezca a la familia conjugada. En este contexto, por el Ejemplo 1.2.1, se tiene

$$\begin{aligned} p(y|\eta) &= b(y,\phi)\exp\{\phi[y\eta - a(\eta)]\}, \qquad \phi > 0 && (1.18)\\ p(\eta) &= c(r,s)\exp\{r\eta - sa(\eta)\}. && (1.19) \end{aligned}$$

Luego (1.16) y (1.17) resultan

$$\begin{aligned} p(\eta|y) &\propto p(y|\eta)p(\eta) \\ &\propto \exp\{(r+\phi y)\eta - (s+\phi)a(\eta)\} \end{aligned}$$

así la distribución de probabilidades a posteriori para η resulta

$$p(\eta|y) = c(r+\phi y, s+\phi)\exp\{(r+\phi y)\eta - (s+\phi)a(\eta)\}. \tag{1.20}$$

Mientras que la distribución de probabilidades para la predicción resulta

$$\begin{aligned} p(y) &= \int p(y|\eta)p(\eta)d\eta \\ &= c(r,s)b(y,\phi)\int \exp\{(r+\phi y)\eta - (s+\phi)a(\eta)\}d\eta \\ &= \frac{c(r,s)b(y,\phi)}{c(r+\phi y, s+\phi)}. \end{aligned} \tag{1.21}$$

Es importante notar que, si la a priori no estuviese en la familia conjugada los cálculos se complicarían considerablemente. En general no sería tratable analíticamente. Así se tendría que recurrir por ejemplo a la integración numérica, o algún metodo de simulación. El problema con estos procedimientos es que muchas veces computacionalmene son muy "caros" de implememtar.

Si denotamos a $\mathbb{E}(Y|\eta)$ por μ y suponemos que $a(\cdot)$ es convexa y dos veces diferenciable, se puede deducir que

$$\begin{aligned} \mathbb{E}(Y|\eta) &= \mu = \dot{a}(\eta) \\ \mathbb{V}(Y|\eta) &= \frac{\ddot{a}(\eta)}{\phi}. \end{aligned}$$

Por estas relaciones, $\dot{a}(\cdot)$ se conoce como la función de media y $\ddot{a}(\cdot)$ como la función de varianza. Bajo estas condiciones $\dot{a}(\cdot)$ es monótona creciente por lo tanto invertible. Esto es, $\eta = \dot{a}^{-1}(\mu)$, de modo que en lugar de η podríamos hacer también todos los análisis en términos de μ.

Por otra parte, a veces es mucho más conveniente trabajar en la escala de μ que en la escala del parámetro natural, como ilustra en el siguiente ejemplo.

1.2.3 Ejemplos

Ejemplo 1.2.1 (Modelo Poisson–Gama) *Para la variable aleatoria Poisson con parámetro de media μ, su a priori conjugada resulta una Gama de parámetros positivos r y s.*

En efecto, sean $Y_1, Y_2, \cdots, Y_n$ iid Poisson(μ), para $\mu > 0$

$$\begin{aligned} p(y_1, y_2, \cdots, y_n|\mu) &= \prod_{i=1}^{n} \frac{\mu^{y_i} e^{-\mu}}{y!} \\ &= \left(\prod_{i=1}^{n} y!\right)^{-1} \exp(-n\mu) \exp\left\{\log \mu \sum_{i=1}^{n} y_i\right\}. \end{aligned} \tag{1.22}$$

Por la aplicación de la Proposición 1.2.1, la densidad a priori conjugada está dado por

$$\begin{aligned} p(\mu|\tau_0, \tau_1) &\propto \exp(-\tau_0\mu) \exp\{\log \mu \ \tau_1\} \\ &= \frac{1}{K(\tau_0, \tau_1)} \mu^{\tau_1} \exp(-\tau_0\mu) \end{aligned}$$

si asumimos la existencia de la integral, para un rango apropiado para los parámetros

$$K(\tau_0, \tau_1) = \int_0^\infty \mu^{\tau_1} \exp(-\tau_0\mu) d\mu$$

se tiene que para $r = \tau_1 + 1$, $s = \tau_0$: $p(\mu|r, s) \propto \mu^{r-1} e^{s\mu}$. Luego, la densidad de la distribución de probabilidades para la a priori asociada a la variable aleatoria Poisson, está dado por

$$p(\mu|\tau_0, \tau_1) = p(\mu|r, s) = Ga(\mu|r, s), \qquad r > 0, \quad s > 0. \tag{1.23}$$

Así, podemos escribir

$$Y|\mu \sim Poisson(\mu), \qquad \mu \sim Gama(r,s).$$

En otras palabras, suponga que se dispone de una observación, digamos y, que puede ser considerada como la realización de una variable aleatoria Y con distribución de probabilidades Poisson de parámetro μ. La función de probabilidades de Y es,

$$p(y) = \frac{e^{-\mu}\mu^y}{y!}, \qquad y = 0, 1, 2, \cdots$$

Ahora suponga que el parámetro μ también es considerado como una variable aleatoria. Así la distribución de probabilidades presentada para Y es condicional en un valor del parámetro μ y es mejor representarla por $p(y|\mu)$.

La densidad de μ con parámetros r y s (ambos positivos) es

$$p(\mu) = \frac{s^r \mu^{r-1} e^{-s\mu}}{\Gamma(r)}, \qquad \mu > 0.$$

La media de μ es (r/s) mientras que la varianza es (r/s^2). Por el Teorema de Bayes, la distribución de probabilidades de μ condicional en la obsevación $Y = y$, es

$$\begin{aligned} p(\mu|y) &\propto p(y|\mu)p(\mu) \\ &= \frac{1}{\Gamma(r)y!}\mu^{r+y-1}e^{-(s+1)\mu}, \qquad \mu > 0. \end{aligned}$$

Naturalmente la variable aleatoria μ tiene distribución a posteriori Gama pero los parámetros r y s se actualizan a $r + y$ y $s + 1$ respectivamente. Si se define la cantidad

$$p = \frac{s}{s+1}.$$

Los momentos posteriores resultan

$$\begin{aligned} \mathbb{E}(\mu|y) &= \frac{r+y}{s+1} = \mathbb{E}(\mu)p + (1-p)y \\ \mathbb{V}(\mu|y) &= \frac{r+y}{(s+1)^2} = (1-p)\mathbb{E}(\mu|y). \end{aligned}$$

Esto quiere decir que el valor esperado de μ condicional en $Y = y$ es un promedio ponderado entre el valor esperado de la distribución a priori y la observación y. El ponderador p depende de la dispersión en la distribución inicial. Mientras mayor el valor del parámetro s, más cercano a 1 es el valor de p y por tanto más importancia se le asigna a la media de la distribución inicial, un valor pequeño de s está relacionado con la alta varianza en la distribución inicial y por tanto la observación y tiene una gran importancia en el cálculo de la media a posteriori. En un caso

extremo, pero realista, en que la información inicial respecto de μ es difusa, se puede utilizar la especificación

$$\frac{r}{s} \mapsto c, \qquad \frac{r}{s^2} \mapsto \infty.$$

Esto implica que $p \mapsto 0$ y por tanto $\mathbb{E}(\mu|y) = \mathbb{V}(\mu|y) = y$.

Por último, una aplicación apropiada de la relación deducida en la Ecuación (1.21), después de algunas simplificaciones resulta que la distribución marginal de Y tiene la forma de una Binomial–Negativa, dada por

$$p(y) = \frac{\Gamma(r+y)}{\Gamma(r)y!} p^r (1-p)^y, \qquad y = 0, 1, 2, \cdots \tag{1.24}$$

donde

$$p = \frac{s}{s+1},$$

y los momentos de esta distribución son

$$\begin{aligned} \mathbb{E}(Y) &= r(1-p)p^{-1} = \frac{r}{s} \\ \mathbb{V}(Y) &= r(1-p)p^{-2} = \frac{r}{s^2}(s+1). \end{aligned}$$

Estos resultados se pueden extender en forma natural al caso en que $Y = y$ representa la suma de n variables aleatorias Poisson.

Ejemplo 1.2.2 (Modelo Binomial–Beta) *En el caso Binomial, su a priori conjugada para el parámetro de probabilidad de éxito μ, es una Beta. En efecto, para un n conocido, sean $Y_1, Y_2, \cdots, Y_m$ iid $Bin(n, \mu)$, $\quad 0 < \mu < 1$*

$$\begin{aligned} p(y_1, y_2, \cdots, y_m|\mu) &= \prod_{i=1}^{m} \binom{n}{y_i} \mu^{y_i} (1-\mu)^{n-y_i} \\ &= \prod_{i=1}^{m} \binom{n}{y_i} \cdot (1-\mu)^{nm} \exp\left\{ \log\left(\frac{\mu}{1-\mu}\right) \sum_{i=1}^{m} y_i \right\}. \end{aligned} \tag{1.25}$$

Por la Proposición 1.2.1, la densidad a priori conjugada está dada por

$$\begin{aligned} p(\mu|\tau_0, \tau_1) &\propto (1-\mu)^{n\tau_0} \exp\left\{ \log\left(\frac{\mu}{1-\mu}\right) \tau_1 \right\} \\ &\propto \mu^{\tau_1} (1-\mu)^{n\tau_0 - \tau_1} \\ &= \frac{1}{K(\tau_0, \tau_1)} \mu^{\tau_1} (1-\mu)^{n\tau_0 - \tau_1} \end{aligned}$$

asumiendo la existencia de la integral

$$K(\tau_0, \tau_1) = \int_0^1 \mu^{\tau_1} (1-\mu)^{n\tau_0 - \tau_1} d\mu$$

Tomamos $r = \tau_1 + 1$, $s = n\tau_0 - \tau_1 + 1$, *con lo cual resulta:* $p(\mu|r,s) \propto \mu^{r-1}(1-\mu)^{s-1}$. *Luego, la densidad de la distribución de probabilidades a priori asociada a la variable aleatoria Binomial, está dado por*

$$p(\mu|\tau_0,\tau_1) = p(\mu|r,s) = Beta(\mu|r,s), \qquad r > 0, \quad s > 0. \tag{1.26}$$

Así, para n es conocido

$$\begin{array}{rcll} Y|\mu & \sim & Bin(n,\mu), & \mu \sim Beta(r,s) \\ p(\mu) & \propto & \mu^{r-1}(1-\mu)^{s-1}, & 0 \le \mu \le 1 \end{array}$$

donde la constante normalizadora es

$$c(r,s) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}. \tag{1.27}$$

La distribución a posteriori para μ dado $Y = y$, que es naturalmente una Beta, que se deduce de una aplicación semejante a la dada por la Ecuación (1.20) como sigue

$$\begin{array}{rcl} p(\mu|y) & \propto & p(y|\mu)p(\mu) \\ & \propto & \mu^y(1-\mu)^{n-y}\mu^{r-1}(1-\mu)^{s-1} \\ & \propto & \mu^{r+y-1}(1-\mu)^{s+n-y-1}. \end{array}$$

Así la distribución a postriori resulta: $p(\mu|y) \propto Beta(r+y, s+n-y)$, y

la distribución predictiva o la marginal de y es una *Beta-Binomial.* En efecto

$$\begin{array}{rcl} p(y) & = & \displaystyle\int_0^1 p(y|\mu)p(\mu)d\mu \\ & = & c(r,s)\begin{pmatrix} n \\ y \end{pmatrix}\displaystyle\int_0^1 \mu^{r+y-1}(1-\mu)^{s+n-y-1}d\mu \\ & = & \begin{pmatrix} n \\ y \end{pmatrix}\dfrac{c(r,s)}{c(r+y,s+n-y)}. \end{array}$$

DIVISION BIENES E INVENTARIOS
VERIFICADO
Fecha
H. M. R
M S

Por lo tanto

$$p(y) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)\Gamma(r+s+n)}\begin{pmatrix} n \\ y \end{pmatrix}\Gamma(r+y)\Gamma(s+n-y) \qquad y = 0,1,2,\ldots,n. \tag{1.28}$$

donde sus dos primeros momentos estan dados por

$$\mathbb{E}(y) = n\frac{r}{r+s}, \qquad \mathbb{V}(y) = \frac{nrs}{(r+s)^2} \times \frac{r+s+n}{r+s+1}. \tag{1.29}$$

En el siguiente ejemplo se considera una extensión del caso Binomial univariado a un proceso multivariado, la cual nos permitirá formular modelos mucho más realistas al incorporar las correlaciones existentes entre las variables en estudio.

Ejemplo 1.2.3 (Modelo Multinomial–Dirichlet) *De una manera natural, los resultados del ejemplo anterior se puede extender para el caso de la variable aleatoria* Multinomial *de dimensión k, cuya a priori conjugada para el vector de probabilidades* $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_k)$, *resulta la variable aleatoria* Dirichlet *de parámetros* $\boldsymbol{r}$, s *y* n, *con* $\boldsymbol{r} = (r_1, r_2, \cdots, r_k)$. *Donde las probabilidades antes mencionadas satisfacen la restricción* $\Sigma\mu_i < 1$. *Esto es, para* n *conocido*

$$\boldsymbol{Y}|\boldsymbol{\mu} \sim Mult(n, \boldsymbol{\mu}), \qquad \boldsymbol{\mu} \sim Dirichlet(\boldsymbol{r}, s)$$

cuya función de probabilidad para un valor observado $\boldsymbol{Y} = \boldsymbol{y}$, *y la densidad para* $\boldsymbol{\mu}$ *estan dados por*

$$Mult_k(\boldsymbol{y}|\boldsymbol{\mu}) = \frac{n!}{\prod_{i=1}^k y_i!(n - \sum_{i=1}^k y_i)!} \prod_{i=1}^k \mu_i^{y_i} (1 - \sum_{i=1}^k \mu_i)^{n - \sum_{i=1}^k y_i}, \tag{1.30}$$

$$Dirich_k(\boldsymbol{\mu}) = \frac{\Gamma(\sum_{i=1}^k r_i + s)}{\prod_{i=1}^k \Gamma(r_i)\Gamma(s)} \mu_1^{r_1 - 1} \cdots \mu_k^{r_k - 1} (1 - \sum_{i=1}^k \mu_i)^{s-1} \tag{1.31}$$

para $y_i = 0, 1, 2, \ldots$, *tal que* $\sum_{i=1}^k y_i \leq n$. *El vector de medias y la matriz de varianzas y covarianzas estan dadas por*

$$\mathbb{E}(\boldsymbol{y}|\boldsymbol{\mu}) = n\boldsymbol{\mu}, \qquad \mathbb{E}(\boldsymbol{\mu}) = \frac{\boldsymbol{r}}{\sum_{i=1}^k r_i + s},$$

$$\mathbb{V}(\boldsymbol{y}|\boldsymbol{\mu}) = n(\text{ diag }(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}'), \qquad \mathbb{V}(\boldsymbol{\mu}) = \frac{\text{diag }(\mathbb{E}(\boldsymbol{\mu})) - \mathbb{E}(\boldsymbol{\mu})\mathbb{E}(\boldsymbol{\mu})'}{\sum_{i=1}^k r_i + s + 1}.$$

De manera análoga, con la aplicación del Teorema de Bayes, la densidad a posteriori de $\boldsymbol{\mu}$ dado la observación $\boldsymbol{Y} = \boldsymbol{y}$, naturalmente es una $Dirich(\boldsymbol{r} + \boldsymbol{y}, s + n - \sum_{i=1}^k y_i)$.

La distribución marginal de $\boldsymbol{Y}$, en este caso resulta una *Dirichlet–Multinomial* cuya densidad está dada por

$$Dir\text{–}Mult(\boldsymbol{y}) = \frac{n!}{\left(\sum_{i=1}^k r_i + s\right)^{[n]}} \prod_{i=1}^k \left(\frac{r_i^{[y_i]}}{y_i!}\right) \frac{s^{[n - \sum_{i=1}^k y_i]}}{(n - \sum_{i=1}^k y_i)!} \tag{1.32}$$

para $y_i = 0, 1, 2, \ldots$, tal que $\sum_{i=1}^k y_i \leq n$, donde $\alpha^{[y]} = \prod_{j=1}^y (\alpha + j - 1)$.

Si definimos $\boldsymbol{p} = \boldsymbol{r}\left(\sum_{i=1}^k r_i + s\right)^{-1}$, los primeros momentos están dados por

$$\mathbb{E}(\boldsymbol{y}) = n\boldsymbol{p}, \qquad \mathbb{V}(\boldsymbol{y}) = \frac{\sum_{i=1}^k r_i + s + n}{\sum_{i=1}^k r_i + s + n + 1} n[\text{ diag }(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}']. \tag{1.33}$$

Por otra parte, la distribución marginal de $(y_1, y_2, \ldots, y_l)$ es una Dirichlet–Multinomial con parámetros $(r_1, r_2, \ldots, r_l, \sum_{j=1}^{k+1} r_j - \sum_{j=1}^l r_j)$ y n, donde $r_{k+1} = s$. Así en particular la distribución marginal de cada y_i es una Beta–Binomial. Más aún, la distribución condicional de $(y_{l+1}, \ldots, y_k)$ dado $(y_1, y_2, \ldots, y_l)$ es también una Dirichlet–Multinomial, con parámetros $(r_{l+1}, \ldots, r_k, \sum_{j=1}^{k+1} r_j - \sum_{j=l+1}^k r_j)$ y $n - \sum_{i=1}^l y_i$.

Ejemplo 1.2.4 (Modelo Normal y Normal–Gama)
Sean $Y_1, Y_2, \cdots, Y_n$ *iid* $N(\mu, \phi)$, $\phi > 0$ *(parámetro de precisión)* y $\mu \in \mathbb{R}$

$$\begin{aligned} p(y_1, y_2, \cdots, y_n|\theta) &= \prod_{i=1}^{n} \left(\frac{\phi}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\phi}{2}(y_i - \mu)^2\right\} \\ &= (2\pi)^{-\frac{n}{2}} \left[\phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}\mu^2\right)\right]^n \exp\left\{\phi\mu \sum_{i=1}^{n} y_i - \frac{\phi}{2}\sum_{i=1}^{n} y_i^2\right\}. \end{aligned} \tag{1.34}$$

Por la Proposición 1.2.1, la densidad a priori conjugada está dada por

$$\begin{aligned} p(\theta|\tau_0, \tau_1, \tau_2) &\propto \left[\phi^{\frac{1}{2}} \exp\left(-\frac{\phi}{2}\mu^2\right)\right]^{\tau_0} \exp\left\{\phi\mu\tau_1 - \frac{\phi}{2}\tau_2\right\} \\ &= \frac{1}{K(\tau_0, \tau_1, \tau_2)} \phi^{\frac{\tau_0 - 1}{2}} \exp\left(-\frac{\phi}{\tau_2}\right) \phi^{\frac{1}{2}} \exp\left\{-\frac{\phi}{2}\tau_0(\mu - \frac{\tau_1}{\tau_0})^2\right\} \end{aligned}$$

asumiendo la existencia de

$$K(\tau_0, \tau_1, \tau_2) = \int_0^\infty \phi^{\frac{\tau_0 - 1}{2}} \exp\left(-\frac{\phi}{\tau_2}\right) \left\{\int_{-\infty}^{\infty} \phi^{\frac{1}{2}} \exp\left[-\frac{\phi}{2}\tau_0\left(\mu - \frac{\tau_1}{\tau_0}\right)^2\right] d\mu\right\} d\phi$$

escribiendo $r = \frac{1}{2}(\tau_1 + 1)$, $s = \frac{1}{2}\tau_2$, $x = \tau_1/\tau_0$ *tenemos*

$$p(\mu, \phi|r, s, x) \propto \phi^{r-1} \exp(-\phi s)\phi^{\frac{1}{2}} \exp\left\{-\frac{\phi(2r-1)}{2}(\mu - x)^2\right\}.$$

Luego

$$p(\theta|\tau_0, \tau_1, \tau_2) = p(\theta|r, s, x) = N(\mu|x, \phi(2r-1))Ga(\phi|r, s), \quad r > \tfrac{1}{2}, \quad s > 0, \quad x \in \mathbb{R}. \tag{1.35}$$

Sea $Y|\mu \sim N(\mu, \phi)$, si se supone la varianza ϕ^{-1} conocida, la a priori conjugada para la media μ resulta también una Normal, digamos $N(x, s)$, donde s denota el parámetro de precisión. La distribución predictiva para Y lo es también Normal $N(x, Q)$, donde $Q = \phi^{-1} + s^{-1}$. Es bien conocido que los momentos posteriores para μ, dada la observación $Y = y$ estan dados por

$$\begin{aligned} \mathbb{E}(\mu|y) &= x + p(y - x), \qquad 0 < p < 1 \\ \mathbb{V}(\mu|y) &= \frac{1}{s + \phi}. \end{aligned}$$

De modo que, $(\mu|y) \sim N(x + p(y - x), s + \phi)$. Donde $p = \phi(s + \phi)^{-1}$. Por otra parte, podemos observar que la media posterior es actualizado como una combinación lineal convexa entre la media a priori y el error de predicción $e = y - x$ ponderada por la cantidad p.

Universidad Mayor de San Andres
Facultad de Ciencias Puras y Naturales
CARRERA DE MATEMATICA
BIBLIOTECA
LA PAZ - BOLIVIA

Capítulo 2

MODELOS DE SERIES DE TIEMPO PARA FAMILIA EXPONENCIAL

2.1 Preliminares

La clase de modelos lineales generalizados que fueron desarrollados por Nelder y Wedderburn(1972), Baker y Nelder(1983), y por McCullagh y Nelder(1989) asume la independencia condicional de las observaciones con distribución común en la familia exponencial. En el contexto de series de tiempo, lo anterior no es aplicable ya que la naturaleza de las observaciones de las series cronológicas es que están correlacionadas y que evolucionan de alguna manera en el tiempo en función de su pasado.

Suponemos que para cada tiempo t, y_t es la realización de una variable aleatoria Y_t con distribución de probabilidades en la familia exponencial bajo las mismas condiciones pedidas en la Sección 1.2.2. Esto es

$$p(y_t|\eta_t, \phi_t) = b(y_t, \phi_t)\exp\{\phi_t[y_t\eta_t - a(\eta_t)]\}, \qquad y_t \in \mathcal{Y} \tag{2.1}$$

donde η_t es el parámetro natural, $\phi_t > 0$ parámetro de precisión (conocido) y $V_t = \phi_t^{-1}$ parámetro de escala.

El conocimiento previo del proceso que genera la variable Y_t en muchas aplicaciones sugieren que el parámetro η_t deba ser considerado aleatorio y con un cierto grado de evolución en el tiempo por las distintas razones asociadas al problema. Por otra parte, los cambios en la componente η_t por lo general no son predecibles, por eso es mejor considerar como una variable aleatoria.

Denotamos por $D_t = \{Y_t, D_{t-1}\}$ al conjunto de información que contiene a las observaciones de la serie y toda la información histórica relevante hasta el instante t. Entonces la distribución de probabilidades de Y_t en la Ecuación (2.1) más explicitamente se puede denotar por $p(y_t|\eta_t, \phi_t, D_{t-1})$ o $p(y_t|\eta_t, D_{t-1})$ ya que ϕ_t en nuestro contexto es un parámetro conocido, o simplemente la denotaremos por $p(y_t|\eta_t)$ sobreentendiendose la dependencia en su pasado. Luego

La distribución a priori de η_t también estará naturalmente en dependencia en términos de la información disponible al tiempo $(t-1)$. Esto es

$$p(\eta_t|D_{t-1}) = c(r_t, s_t)\exp\{r_t\eta_t - s_t a(\eta_t)\}, \qquad s_t > 0. \tag{2.2}$$

Así solo con el conocimiento de los parámetros r_t y s_t la distribución de probabilidades a priori de η_t queda completamente especificada y la distribución de probabilidades para la predicción a 1-paso descrita en la Sección 1.2.2, ahora queda

$$p(Y_t|D_{t-1}) = \int p(Y_t|\eta_t)p(\eta_t|D_{t-1})d\eta_t = \frac{c(r_t, s_t)b(Y_t, \phi_t)}{c(r_t + \phi_t Y_t, s_t + \phi_t)} \tag{2.3}$$

y la actualización posterior para η_t, conociendo la observación y_t al tiempo t, desarrollada en la Sección citada anteriormente queda de la siguiente forma

$$p(\eta_t|D_t) = c(r_t + \phi_t Y_t, s_t + \phi_t)\exp\{(r_t + \phi_t Y_t)\eta_t - (s_t + \phi_t)a(\eta_t)\}. \tag{2.4}$$

2.2 Especificación del Modelo

Un análisis obvio en este estudio es cómo se obtiene la información respecto de la distribución a priori de η_t. Sin embargo, para formar esta distribución se podría utilizar la información disponible relevante hasta el tiempo $(t-1)$. En otras palabras, se supone la existencia de un grado de dependencia a través del tiempo. Sin embargo, en este trabajo, por razones de resolubilidad analítica de los cálculos involucrados con integración, una vez conocida la distribución de probabilidades de las observaciones, tomaremos la a priori como su distribución a priori conjugada.

Por otra parte, es importante notar que en algunas aplicaciones puede ser mucho más conveniente trabajar con un a priori que no necesariamente sea su conjugado, de hecho no hay razones convincentes para justificar solamente el uso de a priori conjugadas, pero los cálculos a los cuales hacíamos referencia antes podrían complicarse considerablemente dependiendo de la naturaleza de la distribución a priori. Así por ejemplo en estos casos se podría llegar incluso a usar métodos numéricos para resolver por ejemplo las integrales involucradas en el cáculo de los momentos, pero en el contexto de series de tiempo tendriamos que aplicar estos procedimientos iterativos por lo menos tantas veces como sea el número de observaciones de la serie, lo cual aún en la actualidad es una limitación por recursos computacionales y algoritmicos.

Por otra parte, también es importante hacer notar que en muchas aplicaciones puede ser mucho más conveniente trabajar en la escala de $\mu_t = \mathbb{E}(Y_t|\eta_t)$ que con el parámetro natural η_t. Más aún, las a priori conjugadas así como se muestra en la Tabla 1.1 se conocen explícitamente para μ_t. De hecho, estas dos componentes estan relacionados univocamente por ejemplo por la función de enlace natural, el mismo que puede ser especificado una vez conocida la distribución de probabilidades de las observaciones.

La especificación del modelo no está completa sin la especificación de la evolución en el tiempo de la componente η_t. En este punto utilizamos las herramientas propuestas para un

modelo estructural de la forma definida por Harvey(1989) y por West y Harrison(1989). En primer lugar la especificación de un modelo estructural está definida para una función de la componente η_t, digamos

$$\lambda_t = g(\eta_t).$$

En este sentido, λ_t sigue un modelo de series de tiempo del mismo tipo que se utiliza para observaciones con distribución Normal. La diferencia aquí es que el modelo es aplicado a una secuencia de valores λ_t que no son observables.

Los modelos estructurales para series de tiempo constituyen una clase de modelos en los cuales las distintas componentes son incluidas en la especificación de manera de capturar explícitamente los aspectos más sobresalientes de la serie cronológica que se está modelando. El modelo más simple es el Modelo Local Constante definido por

$$\begin{aligned} \lambda_t &= \mu_t \\ \mu_t &= \mu_{t-1} + \omega_t \end{aligned}$$

donde ω_t es un ruido blanco, por lo tanto μ_t sigue un proceso de camino aleatorio y representa el nivel de la tendencia en la escala transformada por la función $g(\cdot)$ y la función de enlace con η_t.

En la práctica no es habitual que las series se ajusten al modelo anterior, frecuentemente las series presentan un comportamiento creciente o decreciente en ciertos períodos. Entonces es necesario incorporar al modelo local constante otra componente que permita hacer evolucionar a la tendencia con una cierta pendiente. Este modelo es llamado Local Lineal, y se escribe para la componente λ_t en la forma

$$\begin{aligned} \lambda_t &= \mu_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t} \end{aligned} \tag{2.5}$$

donde μ_t es el nivel de la tendencia y β_t su pendiente. Las componentes ω_{1t} y ω_{2t} son variables aleatorias ruidos blancos no correlacionados en el tiempo. Por esta razón, β_t evoluciona de acuerdo a un camino aleatorio y el nivel de la tendencia μ_t evoluciona de acuerdo a la tendencia y una pendiente al tiempo $(t-1)$ más un proceso totalmente aleatorio no predecible y se espera que su efecto no perdure en el tiempo.

Finalmente, un modelo más realista para series de tiempo que tienen comportamientos periódicos durante cada cierto tiempo, es el llamado Modelo Estructural Básico que incorpora al modelo anterior la componente *estacional*, que se escribe en la forma

$$\begin{aligned} \lambda_t &= \mu_t + \gamma_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t} \\ \sum_{j=0}^{s-1} \gamma_{t-j} &= \omega_t \end{aligned}$$

donde nuevamente ω_t es un ruido blanco y γ_t representa la componente estacional que evoluciona periodicamente cada s unidades de tiempo, salvo un proceso totalmente aleatorio. Para extender más este modelo, se pueden incorporar los efectos cíclicos u otrs componentes relevantes para la serie en estudio. Una discusión más completa de esta clase de modelos para observaciones con distribución Normal se puede encontrar en Harvey(1989).

Un aspecto importante al considerar las distintas componentes que intervienen en el modelo es que la distribución de probabilidades de las componentes ω_{1t} y ω_{2t} no está especificada, solamente estan especificadas los dos primeros momentos de estas variables aleatorias, lo mismo acontece con las componentes μ_t y β_t. No se ha considerado ninguna distribución de probabilidades para estas variables aleatorias. En la especificación del modelo las únicas variables que tienen una distribución de probabilidades explícita son las observaciones Y_t y el parámetro asociado η_t.

Para efectos de estimación de las distintas componentes de interés en una serie de tiempo, es necesario escribir el modelo en su forma de estado asociado de modo que podamos usar con las modificaciones obvias las recursiones de filtro de Kalman apropiadas para nuestro caso.

Definición 2.2.1 *Para cada tiempo t y para algún n*

- *θ_t es el vector de estado n-dimensional.*
- *F_t conocido, vector regresor n-dimensional.*
- *G_t conocido, matriz de evolución de dimensión $n \times n$.*
- *ω_t vector de evolución de errores n-dimencional de media $\mathbf{0}$ y matriz de varianzas W_t. Es decir*

$$\omega_t \sim [\mathbf{0}, W_t]. \tag{2.6}$$

- *$\lambda_t = F_t'\theta_t$ una función lineal del vector de estado de los parámetros.*
- *$g(\eta_t)$ función real conocida, contínua y monótona.*

Entonces el modelo lineal dinámico generalizado (DGLM) para la serie $\{Y_t\}$ en su forma de estado se define por

$$\begin{array}{ll} \textit{Modelo observacional} & p(Y_t|\eta_t) \in FE \\ \textit{Función de enlace} & g(\eta_t) = \lambda_t \\ \textit{Ecuación de medida} & \lambda_t = F_t'\theta_t \\ \textit{Ecuación de transición} & \theta_t = G_t\theta_{t-1} + \omega_t \end{array} \tag{2.7}$$

donde $\omega_t \sim [\mathbf{0}, W_t]$. La ecuación subyacente de transición es llamada en la literatura ecuación de evolución.

De esta definición se desprende que el caso estándar de modelos lineales generalizado estático es un caso particular, ya que para ello basta tomar $\theta_t = \theta, \forall t$, la matriz de evolución igual a la identidad y la matriz de varianzas $W_t = \mathbf{0}$.

2.2 Estimación

Para la estimación de las distintas componentes del vector de estado, supondremos que la matriz de varianzas y covarianzas W_t es conocida, o estimada con algún procedimiento estadístico estándar, como por ejemplo por máxima verosimilitud si éste depende de algunos parámetros, en nuestro contexto supondremos que W_t evoluciona mediante un factor de descuento respecto a la varianza de entrada del filtro de Kalman, y también como es sabido que en este filtro deben estar especificadas todos los hiperparámetros.

Las distintas componentes de interés estan en el vector de estado θ_t el cual queremos estimar para cada tiempo t ya sea usando la información histórica, la información presente o toda la información disponible, en términos de sus primeros dos momentos, el algoritmo que permite hacer estas estimaciones y cálculos corresponde a una modificación de lo que se conoce como el filtro de Kalman, el cual presentamos a continuación en las siguientes líneas. Una versión más detallada se puede encontrar explorando las páginas de West y Harrison(1989).

Filtro de Kalman

Información Inicial o a Posteriori Inicial

Supongamos que antes de conocer la primera observación en el tiempo $t = 0$, se conoce la media y la varianza para el vector de estado θ_0, este es llamado información inicial, de modo que $(\theta_0|D_0) \sim [m_0, C_0]$, donde

$$\mathbb{E}(\theta_0|D_0) = m_0 \qquad y \qquad \mathbb{V}(\theta_0|D_0) = C_0. \tag{2.8}$$

Es importante, mencionar que, no se conoce necesariamente la distribución de probabilidades de esta componente, solo consideramos en el algoritmo los dos primeros momentos de θ_0. Ahora la ecuación de transición permite obtener la media y la varianza de θ_1 con información al tiempo $t = 0$.

Con la misma idea anterior, supongamos que estamos en un instante t y que disponemos de la información al tiempo $(t-1)$. Es decir, conocemos la media y la varianza posterior de $(\theta_{t-1}|D_{t-1})$ denotado por m_{t-1} y C_{t-1} respectivamente. Así

$$(\theta_{t-1}|D_{t-1}) \sim [m_{t-1}, C_{t-1}]. \tag{2.9}$$

Paso 1: A Priori para λ_t

Ahora con la ecuación de transición (2.7) se sigue que los momentos a priori de θ_t condicional en la información histórica son:

$$\begin{aligned} a_t &= G_t m_t \\ R_t &= G_t C_{t-1} G_t' + W_t \end{aligned} \tag{2.10}$$

de modo que $(\theta_t|D_{t-1}) \sim [a_t, R_t]$ y $W_t = (\frac{1}{\delta} - 1)C_{t-1}$, en este caso δ es llamado el factor de descuento y su valor casi siempre está cerca de 1. Sin embargo, dependiento de la naturaleza de las componentes del vector de estado en algunos casos es conveniente usar más de un factor de descuento y en ese caso la matrix W_t podría resultar diagonal por bloques asociados a cada factor de descuento. En la práctica, especificar la matriz de varianza para el error en general es complejo; hay diversos estudios dedicados a esto pero son poco conocidos En el caso de normalidad por ejemplo esta varianza se considera constante en el tiempo.

Entonces, siempre condicional en la información disponible al tiempo $(t-1)$, ocupando la ecuación de medida (2.7) se puede obtener la media y la varianza de la componente λ_t. Esto es

$$\begin{aligned} f_t &= \mathbb{E}(\lambda_t|D_{t-1}) = F_t' a_t \\ q_t &= \mathbb{V}(\lambda_t|D_{t-1}) = F_t' R_t F_t. \end{aligned} \tag{2.11}$$

Más aún, dada la relación de la ecuación de medida, podemos escribir la distribución conjunta a priori entre λ_t y θ_t en términos de sus momentos a priori. Esto es,

$$\begin{pmatrix} \lambda_t \\ \theta_t \end{pmatrix} \sim \left[\begin{pmatrix} f_t \\ a_t \end{pmatrix}, \begin{pmatrix} q_t & F_t' R_t \\ R_t F_t & R_t \end{pmatrix} \right]. \tag{2.12}$$

Paso 2: Especificación de la Distribución a Priori

Dado que

$$\begin{aligned} \mathbb{E}(g(\eta_t)|D_{t-1}) &= \mathbb{E}(\lambda_t|D_{t-1}) \\ \mathbb{V}(g(\eta_t)|D_{t-1}) &= \mathbb{V}(\lambda_t|D_{t-1}) \end{aligned}$$

los parámetros de la distribución a priori conjugada satisfacen el sistema

$$\begin{aligned} \mathbb{E}(g(\eta_t)|D_{t-1}) &= f_t \\ \mathbb{V}(g(\eta_t)|D_{t-1}) &= q_t. \end{aligned} \tag{2.13}$$

De manera que, para encontrar los valores de r_t y de s_t se debe resolver este sistema de ecuaciones, con el cual la distribución de probabilidades a priori quedará completamente determinada. Sin embargo, aunque en su presentación parece simple, este sistema en general resulta no lineal dependiendo por su puesto de la naturalea de la función de enlace g, lo cual dificulta considerablemente su resolución exacta. En algunos casos se tiene que recurrir a los ya conocidos métodos numéricos o simplemente usar alguna aproximación razonable. Algunas pautas, son mostradas en los ejemplos al final de este capítulo.

Paso 3: La Distribución a Posteriori para η_t

Al tiempo t, conocemos la observación y_t con lo cual como en la Sección 2.1, por el Teorema de Bayes, la distribución posterior para η_t queda

$$p(\eta_t|D_t) = c(r_t + \phi_t y_t, s_t + \phi_t) \exp\{(r_t + \phi_t y_t)\eta_t - (s_t + \phi_t)a(\eta_t)\}. \tag{2.14}$$

Este procedimiento para la actualización para η_t mediante el uso de su a priori conjugada se conoce con el nombre de análisis conjugado. Por otro lado esta distribución de probabilidades posterior muestra también el grado de incertidumbre respecto de η_t.

Ahora, la relación $\lambda_t = g(\eta_t)$, especificada en el modelo nos permite encontrar los momentos posteriores para λ_t. Esto es

$$\begin{aligned} f_t^* &= \mathbb{E}(g(\eta_t)|D_t) \\ q_t^* &= \mathbb{V}(g(\eta_t)|D_t). \end{aligned} \tag{2.15}$$

Paso 4a: Estructura Condicional de $(\theta_t|\lambda_t, D_{t-1})$

Finalmente para cerrar el ciclo, obtenemos los momentos posteriores para el vector del estado. Para esto, inicialmente por un momento suponemos conocido λ_t y calculamos los dos primeros momentos de $(\theta_t|\lambda_t, D_{t-1})$, para ello consideramos primero su estructura condicional. Por el teorema de Bayes

$$\begin{aligned} p(\lambda_t, \theta_t|D_t) &\propto p(\lambda_t, \theta_t|D_{t-1})p(Y_t|\lambda_t) \\ &\propto [p(\theta_t|\lambda_t, D_{t-1})p(\lambda_t|D_{t-1})]p(Y_t|\lambda_t) \\ &\propto p(\theta_t|\lambda_t, D_{t-1})[p(\lambda_t|D_{t-1})p(Y_t|\lambda_t)] \\ &\propto p(\theta_t|\lambda_t, D_{t-1})p(\lambda_t|D_t). \end{aligned}$$

Luego se sigue que

$$p(\theta_t|D_t) = \int p(\theta_t|\lambda_t, D_{t-1})p(\lambda_t|D_t)d\lambda_t, \tag{2.16}$$

donde $p(\lambda_t|D_t)$ se puede obtener directamente de la conjugada posterior de η_t mientras que $p(\theta_t|\lambda_t, D_{t-1})$ que es la a priori condicional lo cual no es conocido completamente, de modo que la integral anterior no se puede resolver. Sin embargo, lo que se puede tener para esta distribución condicional son sus primeros dos momentos siguiendo la idea del mejor predictor lineal, teniendo en cuenta que se conoce los momentos de la conjunta que son necesarios para este propósito.

Paso 4b: Momentos a Posteriori para θ_t

De la Ecuación (2.12), los momentos condicionales pueden ser estimados siguiendo la metodología de estimación del mejor predictor lineal en el sentido que minimiza el error cuadrático medio, el mismo que corresponde al análisis usual en el filtro de Kalman.

$$\begin{aligned} \hat{\mathbb{E}}[\theta_t|\lambda_t, D_{t-1}] &= a_t + R_t F_t(\lambda_t - f_t)/q_t \\ \hat{\mathbb{V}}[\theta_t|\lambda_t, D_{t-1}] &= R_t - R_t F_t F_t' R_t/q_t \end{aligned} \tag{2.17}$$

para cualquier λ_t. Es sabido que en el caso de normalidad de las observaciones, estas estimaciones coinciden con los verdaderos valores.

Examinemos ahora derivamos los resultados incondicionales para θ_t, puesto que en definitiva realmente no observamos λ_t. Con la ayuda de la Ecuación (2.16) se puede deducir las siguientes

relaciones que expresan los momentos posteriores para θ_t en función de los momentos condicionales en λ_t

$$\begin{aligned} \mathbb{E}[\theta_t|D_t] &= \mathbb{E}\{\mathbb{E}[\theta_t|\lambda_t, D_{t-1}]|D_t\} \\ \mathbb{V}[\theta_t|D_t] &= \mathbb{V}\{\mathbb{E}[\theta_t|\lambda_t, D_{t-1}]|D_t\} + \mathbb{E}\{\mathbb{V}[\theta_t|\lambda_t, D_{t-1}]|D_t\}. \end{aligned} \tag{2.18}$$

Estos momentos posteriores pueden ser estimados por la sustitución de los momentos condicionales estimados en la Ecuación (2.17). Luego las expresiones finales para los momentos están dadas en las siguientes líneas, de manera que

$$(\theta_t|D_t) \sim [m_t, C_t].$$

Donde

$$\begin{aligned} m_t &= \mathbb{E}[\hat{\mathbb{E}}\{\theta_t|\lambda_t, D_{t-1}\}|D_t] \\ &= \mathbb{E}(a_t + R_t F_t(\lambda_t - f_t)/q_t|D_t) \\ &= a_t + R_t F_t(\mathbb{E}(\lambda_t|D_t) - f_t)/q_t \\ &= a_t + R_t F_t(f_t^* - f_t)/q_t, \end{aligned} \tag{2.19}$$

$$\begin{aligned} C_t &= \mathbb{V}[\hat{\mathbb{E}}\{\theta_t|\lambda_t, D_{t-1}\}|D_t] + \mathbb{E}[\hat{\mathbb{V}}\{\theta_t|\lambda_t, D_{t-1}\}|D_t] \\ &= \mathbb{V}(a_t + R_t F_t(\lambda_t - f_t)/q_t|D_t) + \mathbb{E}(R_t - R_t F_t F_t' R_t/q_t|D_t) \\ &= R_t F_t F_t' R_t \mathbb{V}(\lambda_t|D_t)/q_t^2 + R_t - R_t F_t F_t' R_t/q_t \\ &= R_t - R_t F_t F_t' R_t(1 - \frac{q_t^*}{q_t})/q_t. \end{aligned} \tag{2.20}$$

Con esto queda completado el ciclo, y estas recursiones deben correr para $t = 1, 2, \cdots, T$. En caso de datos faltantes se recurre a la predicción que desarrollamos más adelante. Al final del tiempo ($t = T$) o para cualquier $t < T$ este algoritmo permite hacer predicciones por lo general en términos de sus probabilidades. Es importante también notar aquí la dependencia de C_t en términos de Y_t a través de q_t^*, lo cual no sucede en el caso de Normalidad de las observaciones, en donde C_t es independiente de los datos.

Por otro lado, la estimación de momentos posteriores de las distintas componentes se conoce como la estimación en línea del modelo que usa la información presente y la histórica, un segundo tipo de estimación que se puede calcular, es la estimación suavizada en la cual los valores esperados y las varianzas son calculados en base al conjunto de toda la información disponible, en otras palabras usa la información inicial que da origen a la especificación inicial m_0 y C_0 y las observaciones $y_1, y_2, \cdots, y_T$.

2.4 Predicción

Suponga que para algún tiempo t, en la iteraciones del algoritmo del filtro de Kalman, en la etapa del análisis conjugado se tienen calculadas los valores de los parámetros de la a priori, siempre

condicional al tiempo $(t-1)$, con el cual queda completamente especificada su distribución y de modo que por la Ecuación (2.3) la distribución de probabilidades de la predicción a 1–paso con información al tiempo $(t-1)$ es

$$p(Y_t|D_{t-1}) = \frac{c(r_t, s_t)b(Y_t, \phi_t)}{c(r_t + \phi_t Y_t, s_t + \phi_t)}, \qquad y_t \in \mathcal{Y}. \tag{2.21}$$

En algunos casos, esta distribución resulta ser una conocida, como es el caso del modelo Binomial, en donde la distribución de probabilidades predictiva es una Beta–Binomial y el caso de un modelo Poisson toma la forma de una variable aleatoria Binomial–Negativa.

También es posible obtener predicciones con horizontes superiores a una unidad de tiempo, ya sea al final de la serie, es decir para el futuro o en cualquier instante del tiempo t, $(0 < t < T)$, especialmente para salvar los datos faltantes lo cual no es permitido en el algoritmo del filtro de Kalman, o simplemente para validar en cierto sentido el modelo. Sin embargo, el procedimiento es análogo al ya descrito anteriormente, solo que en lugar de tener la información hasta el tiempo $(t-1)$ se tendrá ahora al tiempo t y la predicción se hará para el tiempo $t+k$, donde k representa el número de pasos del horizonte. Es decir, tenemos las estimaciones posteriores para los momentos de θ_t

$$\theta_t|D_t \sim [m_t, C_t]. \tag{2.22}$$

En este sentido para la predicción a $k-pasos$, aplicamos recursivamente la ecuación de evolución en los tiempos $t+1, t+2, \cdots, t+k$. Luego se sigue que $(\theta_{t+k}|D_t) \sim [a_t(k), R_t(k)]$, con los momentos resursivamente definidos vía

$$\begin{aligned} a_t(k) &= G_{t+k}a_t(k-1) \\ R_t(k) &= G_{t+k}R_t(k-1)G'_{t+k} + W_{t+k} \end{aligned} \tag{2.23}$$

para $k = 1, 2, \cdots$, con $a_t(0) = m_t$, $R_t(0) = C_t$ y $W_{t+k} = W_t$, $\forall k$. Ahora de la relación $\lambda_{t+k} = F'_{t+k}\theta_{t+k}$, tenemos que $(\lambda_{t+k}|D_t) \sim [f_t(k), q_t(k)]$, donde

$$\begin{aligned} f_t(k) &= F'_{t+k}a_t(k) \\ q_t(k) &= F'_{t+k}R_t(k)F_{t+k}. \end{aligned} \tag{2.24}$$

Recordemos que la a priori conjugada para η_{t+k} queda determinada por la espeficación de sus parámetros $r_t(k)$ y $s_t(k)$ los cuales de manera análoga como antes se determinan del sistema

$$\begin{aligned} \mathbb{E}(\lambda_{t+k}|D_t) &= f_t(k) \\ \mathbb{V}(\lambda_{t+k}|D_t) &= q_t(k) \end{aligned} \tag{2.25}$$

con el cual, de acuerdo a la relación presentada en la Ecuación (2.3), la distribución de probabilidades de la predicción a $k-pasos$ está dada por

$$p(Y_{t+k}|D_t) = \frac{c(r_t(k), s_t(k))b(Y_{t+k}, \phi_{t+k})}{c(r_t(k) + \phi_{t+k}Y_{t+k}, s_t(k) + \phi_{t+k})}, \qquad Y_{t+k} \in \mathcal{Y}. \tag{2.26}$$

En los siguientes ejemplos presentamos las partes operacionales que ilustran a los pasos 3 y 4 del filtro de Kalman, puesto que los otros pasos son mera aplicación de las fórmulas.

2.5 Ejemplos

Ejemplo 2.5.1 (Modelo Binomial: Enlace Identidad) *En este modelo tomamos:* $\lambda_t = \mu_t = F_t'\theta_t$, $g^{-1} = \text{logit}$. *Para* n_t *conocido*

$$Y_t|\mu_t \sim Binomial(n_t, \mu_t), \qquad \mu_t|D_{t-1} \sim Beta(r_t, s_t). \tag{2.27}$$

La media y la varianza a priori en este caso estan dados por

$$f_t = \mathbb{E}(\mu_t|D_{t-1}) = \frac{r_t}{r_t + s_t}, \qquad q_t = \mathbb{V}(\mu_t|D_{t-1}) = \frac{f_t(1-f_t)}{r_t + s_t + 1} \tag{2.28}$$

cuyas soluciones para r_t y s_t están dadas por

$$r_t = f_t[q_t^{-1}f_t(1-f_t) - 1], \qquad s_t = (1-f_t)[q_t^{-1}f_t(1-f_t) - 1] \tag{2.29}$$

y los momentos posteriores, una vez conocida la observación y_t, resultan

$$f_t^* = \mathbb{E}(\mu_t|D_t) = \frac{r_t + y_t}{r_t + s_t + n_t}, \qquad q_t^* = \mathbb{V}(\mu_t|D_t) = \frac{f_t^*(1-f_t^*)}{r_t + s_t + n_t + 1}. \tag{2.30}$$

Ejemplo 2.5.2 (Modelo Binomial: Enlace logístico) *Nuevamente consideramos datos Binomiales, bajo las mismas condiciones del ejemplo anterior, la diferencia ahora es que trabajaremos en la escala del parámetro natural* $\eta_t = \text{logit}(\mu_t)$, *la función* $g(\cdot)$ *es la identidad. En cierta manera se complica levemente los procedimientos ya que la a priori si bien es cierto que podemos conocer su distribución de probabilidades, pero en este caso no resulta una de las comunes de los cuales se conocen por fórmula la expresión de sus momentos; sin embargo, haciendo cálculos auxiliares aún es posible explicitar analíticamente la expresión de sus momentos. En este caso el modelo resulta*

$$\lambda_t = \eta_t = F_t'\theta_t. \tag{2.31}$$

La a priori cunjugada en la escala de μ_t *es* $(\mu_t|D_{t-1}) \sim Beta(r_t, s_t)$ *con densidad*

$$p(\mu_t|D_{t-1}) = c(r_t, s_t)\mu_t^{r_t-1}(1-\mu_t)^{s_t-1}, \qquad (0 \le \mu_t \le 1) \tag{2.32}$$

donde

$$c(r_t, s_t) = \frac{\Gamma(r_t + s_t)}{\Gamma(r_t)\Gamma(s_t)}.$$

La media y la varianza a priori de η_t, con la ayuda de los cálculos adicionales desarrollados en el Apéndice A estan dados por

$$\begin{array}{rclcl} f_t & = & \mathbb{E}(\eta_t|D_{t-1}) & = & \gamma(r_t) - \gamma(s_t) \\ q_t & = & \mathbb{V}(\eta_t|D_{t-1}) & = & \dot{\gamma}(r_t) + \dot{\gamma}(s_t). \end{array} \tag{2.33}$$

Donde, para $x > 0$, $\gamma(x) = \dot{\Gamma}(x)/\Gamma(x)$ llamada en la literatura la función *digamma* o la función *psi*, en cuyo caso se denota por $\psi(x)$, en otras palabras esta función es exactamente la derivada

de la función $\log \Gamma(x)$.

Del sistema (2.33), se debe obtener los valores de r_t y s_t, lo cual no es inmediato por la complejidad de la función digamma. Tendríamos que pensar en alguna aproximación razonable[1] hasta considerar métodos de resolución por medio de algoritmos iterativos como por ejemplo el método de Newton–Rapson.

Para valores de r_t y s_t moderadamente grandes (excediendo ambas 3) la a priori para μ_t está considerablemente concentrada lejos de 0 y 1, en este caso la función digamma puede ser adecuadamente aproximado por

$$\gamma(x) = \log x - \frac{1}{2x}, \tag{2.34}$$

la correspondiente aproximación para la derivada es

$$\dot{\gamma}(x) = \frac{1}{x}\left(1 + \frac{1}{2x}\right). \tag{2.35}$$

Sin embargo el sistema aún es complicado de resolver analíticamente.

Para valores grandes de x, $\gamma(x) \approx \log x$, y $\dot{\gamma}(x) \approx 1/x$. Así para valores grandes de r_t y s_t la media y la varianza a priori para η_t aproximadamente estan dados por

$$f_t \approx \log\left(\frac{r_t}{s_t}\right), \qquad q_t \approx \frac{1}{r_t} + \frac{1}{s_t}. \tag{2.36}$$

En este caso particular los valores de r_t y s_t estarían dados aproximadamente por las siguientes expresiones

$$r_t \approx q_t^{-1}(1 + e^{f_t}), \qquad s_t \approx q_t^{-1}(1 + e^{-f_t}). \tag{2.37}$$

De cualquier modo, una vez resuelto el sistema (2.33) los momentos posteriores para η_t son relativamente más sencillos de evaluar. Estos están dados por

$$\begin{array}{rclcl} f_t^* & = & \mathbb{E}(\eta_t|D_t) & = & \gamma(r_t + y_t) - \gamma(s_t + n_t - y_t) \\ q_t^* & = & \mathbb{V}(\eta_t|D_t) & = & \dot{\gamma}(r_t + y_t) + \dot{\gamma}(s_t + n_t - y_t) \end{array} \tag{2.38}$$

puesto que $(\mu_t|D_t) \sim Beta(r_t + Y_t, s_t + n_t - Y_t)$.

En el siguiente ejemplo consideramos la generalización multivariada del modelo Binomial, el cual nos permitirá modelar conjuntamente varias series del mismo tipo.

Ejemplo 2.5.3 (Multinomial) *Sea $\boldsymbol{Y}_t \sim Mult(n_t, \boldsymbol{\mu}_t)$ bajo las mismas condiciones que en el Ejemplo 1.2.3 del Capítulo 1, donde su a priori conjugada para el vector de probabilidades $\boldsymbol{\mu}_t$ es una $Dirich(\boldsymbol{r}_t, s_t)$. Es decir, para n_t conocido*

$$\boldsymbol{Y}_t|\boldsymbol{\mu} \sim Mult(n_t, \boldsymbol{\mu}_t), \qquad \boldsymbol{\mu}_t|D_{t-1} \sim Dirich(\boldsymbol{r}_t, s_t).$$

[1]Este sistema también puede ser resuelto considerando las ecuaciones de recursión presentadas en Abramowitz y Stegun(1965): $\gamma(x) = \gamma(x+1) - x^{-1}$, $\dot{\gamma}(x) = \dot{\gamma}(x+1) + x^{-2}$

El modelo para $\boldsymbol{\lambda}_t$ para este tipo de datos se puede escribir de la siguiente manera

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu}_t \qquad o \qquad \boldsymbol{\lambda}_t = \text{logit}(\boldsymbol{\mu}_t)$$

donde $\text{logit}(\boldsymbol{\mu}_t)$ *se entiende como la función* logit *aplicada a cada componente de* $\boldsymbol{\mu}_t$.

Para el primer caso, los momentos a priori para $\boldsymbol{\lambda}_t$ estan dados por

$$\begin{aligned} \boldsymbol{f}_t &= \mathbb{E}(\boldsymbol{\mu}_t|D_{t-1}) = \frac{\boldsymbol{r}_t}{\sum r_{jt} + s_t} \\ \boldsymbol{Q}_t &= \mathbb{V}(\boldsymbol{\mu}_t|D_{t-1}) = \frac{\text{diag}\,(\boldsymbol{f}_t) - \boldsymbol{f}_t\boldsymbol{f}_t'}{\sum r_{jt} + s_t + 1}. \end{aligned}$$

Resolviendo éste sistema para $\boldsymbol{r}_t$ y s_t se tienen especificados los parámetros para la distribución a priori.

La distribución a posteriori para $\boldsymbol{\mu}_t$ es

$$\boldsymbol{\mu}_t|D_t \sim Dirich(\boldsymbol{r}_t + \boldsymbol{y}_t, s_t + n_t - \sum_{j=1}^{k} y_{jt})$$

En consecuencia, los momentos posteriores estan dados por

$$\begin{aligned} \boldsymbol{f}_t^* &= \mathbb{E}(\boldsymbol{\mu}_t|D_t) = \frac{\boldsymbol{r}_t + \boldsymbol{y}_t}{\sum r_{jt} + s_t + n_t} \\ \boldsymbol{Q}_t^* &= \mathbb{V}(\boldsymbol{\mu}_t|D_t) = \frac{\text{diag}\,(\boldsymbol{f}_t^*) - \boldsymbol{f}_t^*\boldsymbol{f}_t^{*\prime}}{\sum r_{jt} + s_t + n_t + 1}. \end{aligned}$$

En cambio, si consideramos la función de enlace logito, dado que las marginales de una variable Dirichlet es también una Dirichlet, entonces la marginal de cada componente de $\boldsymbol{\mu}_t$ resulta una Beta. Es decir

$$\mu_{jt} \sim Beta(r_{jt}, \sum_{i \neq j}^{k} r_{it} + s_t), \qquad j = 1, 2, \ldots, k.$$

Donde la medias a priori para λ_t, siguiendo los resultados de los ejemplos univariados resulta

$$f_{jt} = \gamma(r_{jt}) - \gamma(\sum_{i \neq j}^{k} r_{jt} + s_t), \qquad j = 1, 2, \ldots, k.$$

donde la función $\gamma(\cdot)$ es la función digamma definido anteriormente. Para obtener la varianza a priori se puede usar la aproximación lineal, más conocido como el método-δ, para la función de enlace logito. Así, tomando la matriz diagonal D de las derivadas de la función de enlace logito con respecto de $\boldsymbol{\mu}_t$ evaluada en $\boldsymbol{f}_t$.

$$D = \text{diag}\left(\frac{1}{f_{1t}(1 - f_{1t})}, \cdots, \frac{1}{f_{kt}(1 - f_{kt})}\right),$$

se tiene,

$$Q_t = DQ_t^+ D$$

donde Q_t^+ es la matriz de varianzas y covarianzas a priori del caso de la función de enlace identidad entre $\boldsymbol{\lambda}_t$ y $\boldsymbol{\mu}_t$.

De la misma manera los momentos posteriores para $\boldsymbol{\lambda}_t$ tienen las siguientes expresiones

$$f_t^* = \gamma(r_t + y_t) - \gamma(\sum r_{it} + r_t + s_t + n_t - y_t)$$

abusando un poco la notación, la suma de un escalar y un vector se entiende como la suma del escalar a cada componente del vector. Para encontrar la matriz de varianzas y covarianzas a posteriori, nuevamente usamos el método-δ. Así Q_t^* resulta

$$Q_t^* = D^* Q_t^{**} D^*$$

Donde Q_t^{**} es la matriz de varianzas y covarianzas para la a posteriori del caso de la función de enlace identidad y

$$D^* = \text{diag}\left(\frac{1}{f_{1t}^*(1-f_{1t}^*)}, \cdots, \frac{1}{f_{kt}^*(1-f_{kt}^*)}\right)$$

Ejemplo 2.5.4 (Modelo Poisson) *Sea Y_t una variable aleatoria* Poisson *de media μ_t, nuevamente consideramos la función de enlace $g(\cdot)$ como la identidad como sucede en varias aplicaciones, luego el modelo queda*

$$\lambda_t = \eta_t = F_t' \theta_t \tag{2.39}$$

La a priori conjugada conocida en la escala de μ_t es una Gama, es decir

$$Y_t|\mu_t \sim Poisson(\mu_t), \qquad \mu_t|D_{t-1} \sim Gama(r_t, s_t). \tag{2.40}$$

Los momentos de la distribución a priori en la escala de μ_t son:

$$\mathbb{E}(\mu_t|D_{t-1}) = \frac{r_t}{s_t}, \qquad \mathbb{V}(\mu_t|D_{t-1}) = \frac{r_t}{s_t^2}. \tag{2.41}$$

Mientras que los momentos de la distribución a priori en la escala de $\eta_t = \log \mu_t$, lo que en este caso realmente nos interesa, se puede encontrar de la manera que está descrita en la primera parte del Apéndice A, cuyos resultados son los siguientes

$$\begin{array}{rclcl} f_t & = & \mathbb{E}(\eta_t|D_{t-1}) & = & \gamma(r_t) - \ln(s_t) \\ q_t & = & \mathbb{V}(\eta_t|D_{t-1}) & = & \dot{\gamma}(r_t). \end{array} \tag{2.42}$$

Aquí otra vez, la resolución de este sistema para r_t y s_t no es inmediato, habría que recurrir nuevamente a algún método numérico o usar una aproximación razonable. Ocupando la aproximación escrita en la Ecuación (2.34) los valores de los parámetros de la distribución a priori resultan

$$r_t = \frac{1 + \sqrt{1 + 2q_t}}{2q_t}, \qquad s_t = r_t \exp\{-f_t - \frac{1}{2r_t}\}. \tag{2.43}$$

Ahora al tiempo t conociendo la observación y_t por el análisis conjugado la posterior para μ_t resulta $(\mu_t|D_t) \sim Gama(r_t + y_t, s_t + 1)$. Luego los momentos posteriores en la escala de η_t se pueden escribir facilmente de la siguiente manera, aunque su evaluación podría resultar otra tarea.

$$\begin{aligned} f_t^* &= \mathbb{E}(\eta_t|D_t) &= \gamma(r_t + y_t) - \ln(s_t + 1) \\ q_t^* &= \mathbb{V}(\eta_t|D_t) &= \dot{\gamma}(r_t + y_t). \end{aligned} \tag{2.44}$$

Estos tres ejemplos ilustran quizás los casos más comunes que se presentan en las aplicaciones.

Ejemplo 2.5.5 (Normal) *Sea $Y_t|\mu_t \sim N(\mu_t, \phi_t)$, donde la varianza ϕ_t^{-1} se supone conocido. Luego su a priori conjugada para μ_t también es una Normal $N(x_t, s_t)$. Si consideramos el modelo $\lambda_t = \mu_t$, los parámetros de la distribución a priori satisfacen el sistema*

$$f_t = \mathbb{E}(\lambda_t|D_t) = x_t, \qquad q_t = \mathbb{V}(\lambda_t|D_t) = s_t^{-1}$$

DIVISIÓN BIENES E INVENTARIOS
VERIFICADO
Fecha
H. M. R.
M. S

la distribución a posteriori para μ_t dada la observación $Y_t = y_t$ es

$$\mu_t|y_t \sim N(x_t + p_t(y_t - x_t), \phi_t + s_t)$$

y la distribución de predicción a 1-paso para Y_t, es también una Normal

$$Y_t|D_{t-1} \sim N(x_t, Q_t),$$

donde la varianza $Q_t = \phi_t^{-1} + s_t^{-1}$. Y finalmente que la varianza a posteriori para μ_t, puede ser escrito también en la forma estandar $s_t^{-1} - p_t^2 Q_t$.

2.6 Suavizamiento

Como se mencionó anteriormente, los estimadores suavizados de la componente del vector de estado se obtienen usando toda la información existente hasta ese momento. Es decir, se quiere estimar los momentos de la distribución de probabilidades de $(\theta_{t-k}|D_t)$ para $k \geq 1$, como en el caso de la predicción del vector de estado, los análogos de $a_t(k)$ y $R_t(k)$ que la denotamos por $a_t(-k)$ y $R_t(-k)$ sus expresiones estan dadas en forma recursiva retrospectiva de la siguiente manera.

Para $1 \leq k < t$, y $\forall t$, definimos

$$B_t = C_t G'_{t+1} R_{t+1}^{-1} \tag{2.45}$$

luego por lo menos, en términos de sus momentos $(\theta_{t-k}|D_t) \sim [a_t(-k), R_t(-k)]$. Donde

$$\begin{aligned} a_t(-k) &= m_{t-k} + B_{t-k}[a_t(-k+1) - a_{t-k+1}] \\ R_t(-k) &= C_{t-k} - B_{t-k}[R_{t-k+1} - R_t(-k+1)]B'_{t-k} \end{aligned} \tag{2.46}$$

los valores iniciales estan dadas por $a_t(0) = m_t$, y $R_t(0) = C_t$. También a_{t-k+1} y R_{t-k+1} se pueden denotar por $a_{t-k}(1)$ y $R_{t-k}(1)$ respectivamente.

Una manera de respaldar estas ecuaciones, es el hecho de que no se asume una distribución de probabilidades para las componentes del vector de estado y solo se trabaje con especificaciones de las medias y las varianzas. Esto hace que el procedimiento para obtener la estimación suavizada de estas componentes sea equivalente al procedimiento usual en el filtro de Kalman; mayores detalles se pueden encontrar en West y Harrison(1989).

2.7 Función de Verosimilitud

La función de verosimilitud para los hiperparámetros del modelo, como son los factores de descuento, se puede escribir recurriendo a la idea de la descomposición del error de predicción usual que se ocupa en series de tiempo, siendo que la distribución para la predicción que se obtiene usando la información a priori y la verosimilitud de las observaciones. Como es usual en la inferencia estadística en lugar de la función de verosimilitud misma, es mejor considerar la log-verosimilitud y observar la cantidad $-2\log L$, ya sea con fines de encontrar la optimalidad o para hacer alguna inferencia relacionada con los hiperparámetros.

$$\ell(hiperparámetros|D_T) = \sum_{t=1}^{T} \log p(y_t|D_{t-1}). \tag{2.47}$$

En general ésta función define implícitamente a los hiperparámetros, lo cual dificulta el uso tradicional de la diferenciación para estimar los parámetros. Más aún, los hiperparámetros están involucrados en las ecuaciones del filtro de Kalman y para poder ser evaluadas una sola vez, se debe correr todo el filtro. Este procedimiento en la literatura estadística se conoce como la estimación en el dominio del tiempo, existiendo también métodos alternativos de estimación.

2.8 Residuo de Pearson

Con la idea desarrollada por Nelder y Wedderburn(1972) para observaciones independientes, el caso de observaciones de una serie de tiempo condicionales en su pasado son aproximadamente no correlacionadas, esto da lugar a lo que en la nomenclatura de series se llamaría residuo de predicción estandarizada o simplemente residuos de Pearson

$$rp_t = \frac{y_t - \mathbb{E}(y_t|D_{t-1})}{\sqrt{\mathbb{V}(y_t|D_{t-1})}}. \tag{2.48}$$

Existiendo además otras formas de definir residuos, pero en nuestro contexto creemos que éste es al más adecuado. Una discusión más detallada con respecto a otros tipo de residuos se puede encontrar en la bibliografía citada anteriormente.

Capítulo 3

APLICACION I: GOLES MARCADOS POR U. CATOLICA EN LOS CLASICOS UNIVERSITARIOS

3.1 Análisis del Problema

En esta primera aplicación, se estudia la evolución dinámica del número de goles marcados por el equipo de la Universidad Católica en partidos del campeonato nacional jugados contra el equipo de la Universidad de Chile desde 1939 y hasta 1994. Los resultados de cada partido se muestran en la Tabla 3.5 al final de este capítulo.

La naturaleza de este tipo de observaciones sugiere que el número de goles marcados en un partido pueda ser considerado como una realización de una variable aleatoria Poisson. Por otra parte, la natural dependencia de las observaciones en el tiempo aconseja el uso de una clase de modelos de serie de tiempo para estudiar y predecir el comportamiento futuro del número de goles marcados. El número de goles marcados en años consecutivos están correlacionados y no pueden ser tratados como realizaciones independientes. Sin embargo, no existen muchos estudios en los cuales se consideren series de tiempo con distribución Poisson.

Para justificar de cierta manera que los datos corresponden a realizaciones de una variable aleatoria Poisson, en la Figura 3.2 se muestra una representación gráfica de los goles marcados por el equipo de la Universidad Católica, y en la Tabla 3.1 se presentan las frecuencias relativas correspondientes a estos goles conjuntamente con las probabilidades teóricas de la variable Poisson condicional en los parámetros cercanos al promedio de goles, cuyos gráficos se puede observar en la Figura 3.1. En este gráfico se puede ver que la aproximación de una Poisson a una Normal claramente no es apropiada, puesto que en ese caso tendríamos que tener el número de goles por partido superior de lo que habitualmente se registra en los partidos de fútbol.

Desde un punto de vista metodológico, los modelos de serie de tiempo que han predominado en la literatura estadística en los últimos años suponen distribución de probabilidades Normal

Tabla 3.1: **Frecuencias Relativas y Probabilidades**

Número de Goles	Frecuencia Relativa	Poisson(1.0)	Poisson(1.5)
0	0.31	0.37	0.22
1	0.31	0.37	0.33
2	0.17	0.18	0.25
3	0.16	0.06	0.13
4	0.03	0.02	0.05
5	0.02	0.00	0.02

para las observaciones, o restringen el análisis a los dos primeros momentos de las variables, y por tanto, a estimaciones y predicciones lineales.

Dos aspectos fundamentales relacionados con los objetivos del estudio tienen que ver con la dependencia temporal de las observaciones y la evolución de los parámetros que determinan la distribución de probabilidades de las observaciones.

En este sentido, la evolución natural que tiene en el tiempo la pericia de cualquier equipo de fútbol hace que las observaciones deban ser consideradas dependientes o correlacioanadas en el tiempo, y el parámetro de las distribuciones Poisson en distintos momentos de tiempo debe evolucionar de acuerdo con alguna especificación.

Luego, considerando al número de goles marcados por un equipo de fútbol como realizaciones de una variable aleatoria Poisson, ajustamos a los datos un modelo dinámico Poisson–Gama que se describe más adelante. El modelo entrega una estimación de la tendencia para el valor esperado de goles y una distribución de probabilidades para el número de goles en el próximo partido.

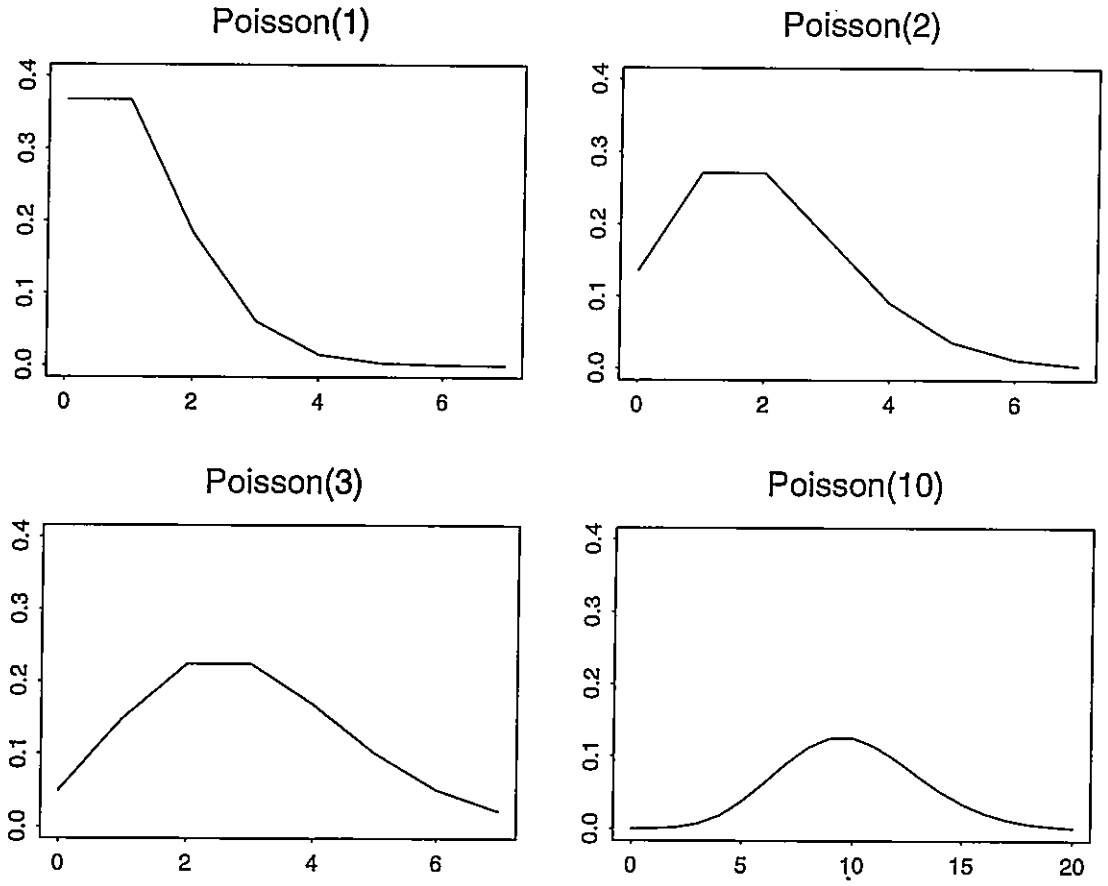
No se conoce muchos otros estudios que consideren problemas similares al abordado en este capítulo. En Harvey y Fernandes(1989) se estudia la evolución del número de goles de Inglaterra a Escocia en un serie de tiempo de más de 50 años. Sin embargo, la metodología propuesta en ese estudio sigue más de cerca la metodología propuesta en Harvey(1989) que la utiliza en este trabajo. El tratamiento que se usa aquí es muy similar al propuesto por West y Harrison(1989) y también por West, Harrison y Migon(1985).

3.2 Un Modelo Dinámico Poisson–Gama

3.2.1 El Número de Goles de la U. Católica

Sea y_t el número de goles marcados por el equipo de la U. Católica en los n_t clásicos universitarios jugados en el año t. Supongamos que y_t es la realización de una variable aleatoria Poisson Y_t de parámetro μ_t, el mismo que representa el número de goles esperado en los n_t partidos del

Figura 3.1: **Probabilidades Poisson**



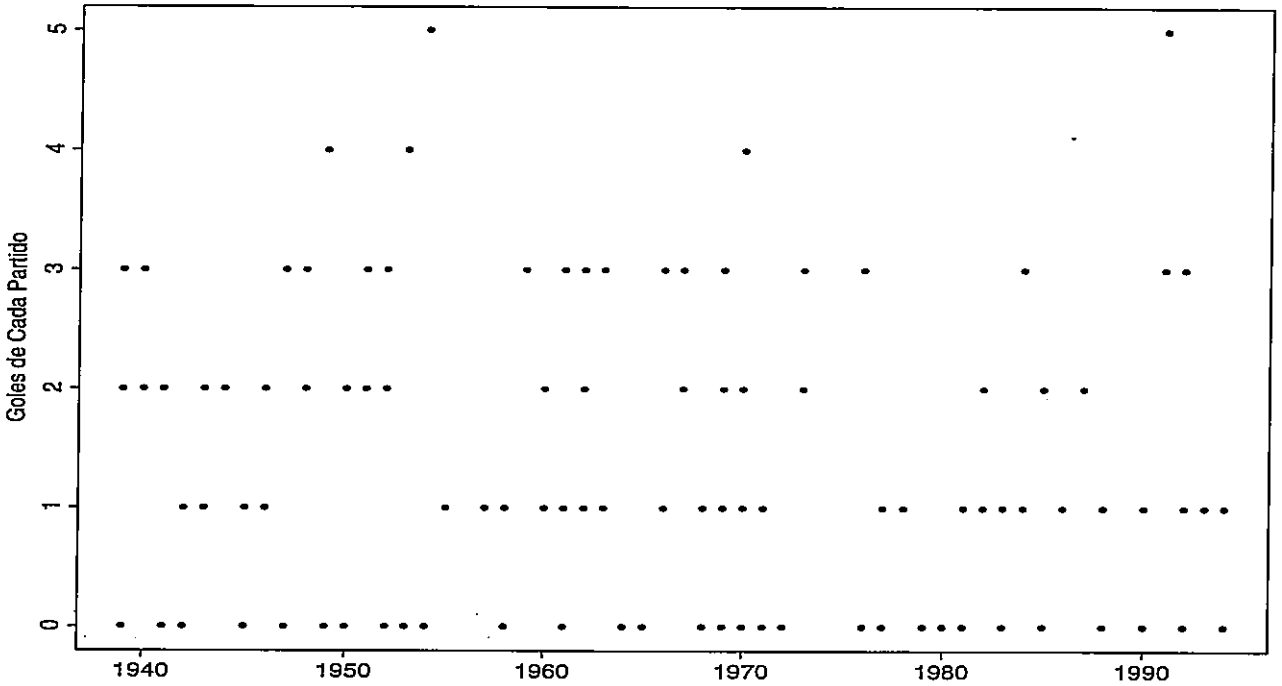
año t. La Figura 3.2 y el conocimiento del proceso que generó la variable Y_t sugiere que μ_t deba ser considerado aleatorio y con un cierto grado de evolución o cambio a través del tiempo. Las razones pueden ser los cambios en la composición del equipo, los cambios en las tácticas de juego, la pericia individual de los jugadores, condiciones sicológicas, etc. Además, como los cambios no son predecibles, es mejor considerar a μ_t como una variable aleatoria.

Así, en términos generales se tiene que para un año t, distintos factores determinan el valor de la variable μ_t. Luego en cada partido jugado en ese año, el número de goles marcados por la Universidad Católica será la realización de una variable aleatoria Poisson con un parámetro igual al valor de μ_t. Posiblemente, un análisis más realista se deba considerar a μ_t cambiando mes a mes o incluso cambiando contínuamente. Para simplificar el análisis y los cálculos aquí se supone que el parámetro μ_t sólo cambia de un año a otro. Para la variable aleatoria μ_t se supone una distribución Gama con parámetros r_t y s_t. Así

$$Y_t|\mu_t \sim Poisson(\mu_t), \qquad \mu_t|D_{t-1} \sim Gama(r_t, s_t), \tag{3.1}$$

donde D_{t-1} es como se definió en la Sección 2.1, que escencialmente contiene la información histórica al tiempo $(t-1)$. Este modelo conocido como Poisson–Gama, requiere que, en cada año

Figura 3.2: Número de Goles de U. Católica



se especifiquen los valores de los parámetros r_t y s_t. Estos coeficientes determinan la distribución de probabilidades a priori para μ_t, que puede ser actualizado al tiempo t, con el conocimiento de la observación $Y_t = y_t$, por los procedimientos descritos en la Sección 1.2.2. En otras palabras, supone la existencia de un grado de dependencia a través del tiempo en el parámetro fundamental μ_t especificada por r_t y s_t que determina la distribución de probabilidades del número de goles.

En este sentido, la especificación del modelo no está completa sin una especificación para la evolución en el tiempo de la componente μ_t. Como se había indicado en el desarrollo metodológico de la Sección 2.2, lo que sigue un modelo estructural es una función de μ_t, digamos $\lambda_t = g(\mu_t)$. Las dos alternativas con las cuales se trabaja en este capítulo para definir la componente λ_t son las siguientes

$$\lambda_t = g(\mu_t) = \begin{cases} \mu_t/n_t \\ \log(\mu_t/n_t). \end{cases} \tag{3.2}$$

La primera de estas especificaciones, es el valor esperado del número de goles en un partido lo que sigue un modelo estructural de series de tiempo, mientras que la segunda de estas especificaciones es el logaritmo natural de este valor esperado el que evoluciona de acuerdo con un modelo estructural de series de tiempo.

En este caso particular, el modelo propuesto para la componente λ_t tiene la forma

$$\begin{aligned} \lambda_t &= \mu_t^* \\ \mu_t^* &= \mu_{t-1}^* + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t} \end{aligned} \tag{3.3}$$

Aquí μ_t^* representa el nivel promedio de goles por partido en la escala original o logarítmica, según sea la función de enlace en la relación (3.2). También se puede interpretar la componente μ_t^* como el indicador de la pericia del equipo. La media μ_t^* cambia en función del nivel del año anterior, en función de la componente β_{t-1}, y en función de una componente complementaria aleatoria y no predecible ω_{1t}. La componente β_t representa a una pendiente dinámica que evoluciona de acuerdo con un camino aleatorio, por cuanto ω_{2t}, como ω_{1t} son componentes completamente aleatorias. Más formalmente, se supone que ω_{2t} y ω_{1t} son variables aleatorias ruido blancos con media cero y no correlacionados en el tiempo, cuya distribución de probabilidades no está especificada.

3.2.2 Estimación

Inicialmente, se supone que las varianzas de ω_{1t} y ω_{2t} son conocidas, y que sigue la idea del factor de descuento descrita en la Sección 2.3. En esta aplicación en particular se usan dos factores de descuento δ_1 y δ_2 funcionalmente independientes asociados a cada componente μ_t^* y β_t respectivamente, y que los mismos serán estimados usando el la idea estándar de máxima verosimilitud.

El algoritmo que permite hacer las estimaciones y los cálculos para las componentes μ_t^* y β_t corresponde a una modificación de lo que se conoce como el filtro de Kalman, lo cual fue descrita anteriormente en el Capítulo 2. Inicialmente para ello, primero escribimos el modelo propuesto en su forma de estado

$$\begin{aligned} \lambda_t &= F'\theta_t \\ \theta_t &= G\theta_{t-1} + \omega_t \end{aligned} \tag{3.4}$$

donde $\omega_t \sim [\mathbf{0}, W_t]$, siendo W_t conocido, y

$$F' = (1,0), \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \omega_t = \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix}, \qquad \theta_t = \begin{pmatrix} \mu_t^* \\ \beta_t \end{pmatrix}. \tag{3.5}$$

Filtro de Kalman

Información Inicial o a Posteriori Inicial

Los valores iniciales m_0 y C_0, que usamos para ajustar este modelo fueron

$$m_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \qquad y \qquad C_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.6}$$

Así

$$(\theta_0|D_0) \sim [m_0, C_0]. \tag{3.7}$$

Paso 1: A Priori para λ_1
Ahora, de la ecuación de transición (3.4), se sigue que los momentos de la distribución a priori de θ_1 condicional en la información histórica son:

$$\begin{aligned} a_1 &= Gm_1 \\ R_1 &= GC_0G' + W_1 \end{aligned} \tag{3.8}$$

de modo que $(\theta_1|D_0) \sim [a_1, R_1]$ y

$$W_1 = \text{diag}\left(\frac{1}{\delta_1} - 1, \frac{1}{\delta_2} - 1\right) \text{diag}\,(C_0) = \text{diag}\,(0.031, 0.053)\ \text{diag}\,(C_0) \tag{3.9}$$

donde los factores de descuento fueron estimados por máxima verosimilitud, la función de verosimilitud se estudia en la Sección (3.2.6), los valores resultaron

$$\hat{\delta}_1 = 0.97 \qquad y \qquad \hat{\delta}_2 = 0.95. \tag{3.10}$$

Luego, siempre condicional en la información disponible al tiempo $t = 0$, ocupando la ecuación de medida (3.4) se puede obtener la media y la varianza de λ_1. Esto es

$$\begin{aligned} f_1 &= \mathbb{E}(\lambda_1|D_0) = F'a_1 \\ q_1 &= \mathbb{V}(\lambda_1|D_0) = F'R_1F. \end{aligned} \tag{3.11}$$

Paso 2: Especificación de la Distribución a Priori
En este caso, los parámetros de la distribución a priori conjugada para el caso $\lambda_t = \mu_t/n_t$ dado en la Ecuación 3.2 satisfacen el sistema

$$\begin{aligned} \mathbb{E}(\mu_1^*|D_0) &= \frac{r_1}{n_1 s_1} = f_1 \\ \mathbb{V}(\mu_1^*|D_0) &= \frac{r_1}{n_1^2 s_1^2} = q_1 \end{aligned}$$

siendo n_1 es conocido, estas ecuaciones definen de manera única los parámetros r_1 y s_1 y por lo tanto la distribución de probabilidades para μ_1 siempre condicional en la información disponible al tiempo $t = 0$.

Por otra parte, si la especificación que relaciona λ_t con μ_t es de la segunda forma propuesta en la Ecuación (3.2), el sistema que satisfacen los parámetros r_1 y s_1, de acuerdo al presentado en el Ejemplo 2.5.4 sería como sigue

$$\begin{aligned} \gamma(r_1) - \ln(s_1) &= f_1 \\ \dot{\gamma}(r_1) &= q_1 \end{aligned} \tag{3.12}$$

como en el caso anterior, estas ecuaciones determinan de manera única los valores de r_1 y s_1 y por tanto la distribución de probabilidades de μ_1 condicional en D_0.

Sin embargo, la resolución de este sistema no es inmediata, aunque algebraicamente su solución se puede escribir como

$$\begin{aligned} r_1 &= \dot{\gamma}^{-1}(q_1) \\ s_1 &= \frac{1}{n_1}\exp\{\gamma(r_1) - f_1\}. \end{aligned}$$

En la implementación con la rutinas computacionales, para invertir $\dot{\gamma}(x)$ usamos el método de Newton-Rapson, disponiendo de una rutina de cálculo de alta precisión para la función digamma y usamos aproximaciones de tipo polinomial para el cáculo de sus derivadas tomando un número considerable de términos de su expansión en serie.

Paso 3: La Distribución a Posteriori para μ_1^*
Al tiempo $t = 1$ conocemos la observación y_1, o la suma del conjunto de observaciones si hay más de un partido, con el cuál de manera análoga como en la Ecuación (2.4), por el Teorema de Bayes, la distribución a posteriori para μ_1 queda

$$\mu_1|D_1 \sim Gama[r_1 + y_1, s_1 + 1] \quad \Longrightarrow \quad \frac{\mu_1}{n_1}\Big| D_1 \sim Gama[r_1 + y_1, n_1(s_1 + 1)]. \tag{3.13}$$

Esta distribución de probabilidades muestra el grado de incertidumbre respecto de la media μ_1. La observación y_1 afecta a ésta distribución en el sentido de que la misma es formada utilizando la información previa al tiempo $t = 0$ y la observación del número de goles al tiempo $t = 1$.

Ahora, la relación $\lambda_t = g(\mu_t)$, especificada en el modelo nos permite encontrar los momentos posteriores para λ_1. Así, para el primer caso resulta

$$\begin{aligned} f_1^* &= \mathbb{E}(\lambda_1|D_1) = \frac{r_1 + y_1}{n_1(s_1 + 1)} \\ q_1^* &= \mathbb{V}(\lambda_1|D_1) = \frac{f_1^*}{n_1(s_1 + 1)}, \end{aligned}$$

mientras que para el modelo con enlace natural log, los momentos posteriores estan dados por

$$\begin{aligned} f_1^* &= \mathbb{E}(\lambda_1|D_1) = \gamma(r_1 + y_1) - \log[n_1(s_1 + 1)] \\ q_1^* &= \mathbb{V}(\lambda_1|D_1) = \dot{\gamma}(r_1 + y_1). \end{aligned} \tag{3.14}$$

Paso 4: Momentos a Posteriori para θ_1
Para terminar el ciclo recursivo y disponer de la media y varianza de las componentes del vector de estado θ_1 condicional en la información $t = 1$. Aplicando las relaciones deducidas en la Sección 2.3, resulta que $(\theta_1|D_1) \sim [m_1, C_1]$. Donde

$$\begin{aligned} m_1 &= \mathbb{E}(\theta_1|D_1) = a_1 + R_1 F(f_1^* - f_1)/q_1 \\ C_1 &= \mathbb{V}(\theta_1|D_1) = R_1 - R_1 F F' R_1 (1 - \frac{q_1^*}{q_1})/q_1. \end{aligned}$$

El procedimiento general propuesto para $t = 0$, se puede repetir en cada momento del tiempo según se obtengan nuevas observaciones. Es decir, los valores iniciales de entrada para un tiempo t son los momentos posteriores al tiempo $(t-1)$. Así se puede obtener en cada momento del tiempo una estimación de λ_t y de μ_t.

3.2.3 Nivel de la Tendencia en Línea y Suavizada

Una vez calculada los parámetros de la distribución a priori de acuerdo a los procedimientos descritos para este caso, la estimación en línea del número esperado de goles por partido (μ_t/n_t) queda determinada por el valor esperado obtenido de la relación en la Ecuación (3.13), cuya representación gráfica se muestra en la Figura. 3.3.

$$\mathbb{E}\left(\frac{\mu_t}{n_t}\middle| D_t\right) = \frac{r_t + y_t}{n_t(s_t+1)}. \tag{3.15}$$

Un procedimiento similar se puede seguir para el caso del nivel suavizado usando toda la información disponible hasta el tiempo t

$$\mathbb{E}\left(\frac{\mu_t(-k)}{n_t(-k)}\middle| D_t\right) = \frac{r_t(-k) + y_{t-k}}{n_{t-k}(s_t(-k)+1)}, \qquad 1 \leq k < t. \tag{3.16}$$

Los resultados de las estimaciones en línea y del nivel suavizado de la tendencia, se muestran

Tabla 3.2: **Estimación de la Tendencia**

Año	Promedio de Goles	Estimación en Línea		Estimación Suavizada	
		$\mathbb{E}[\mu_t/n_t]$	$\mathbb{V}[\mu_t/n_t]$	$\mathbb{E}[\mu_t/n_t]$	$\mathbb{V}[\mu_t/n_t]$
1940	2.50	1.84	0.53	1.78	0.41
1950	1.00	1.48	0.23	1.59	0.14
1960	1.50	1.69	0.22	1.46	0.08
1970	1.75	1.23	0.06	1.17	0.04
1980	0.00	0.70	0.05	0.95	0.05
1990	0.50	0.84	0.06	1.13	0.05

en las Figuras 3.3 y 3.4, de los cuales se desprende que el número promedio de goles por partido marcados por el equipo de la Universidad Católica disminuyó considerablemente aproximadamente desde 1954 y hasta 1980. Este promedio era cercano a 1.60 en 1954 y cercano a 0.95 en 1980. desde 1981 se aprecia una recuperación en el promedio de goles por partido. En 1994 este valor fluctuaba en torno a 1.15. En este trabajo sólo se presentan los resultados para el caso en que λ_t corresponde al logaritmo del número de goles esperado por partido y sólo en el caso de residuos presentamos la gráfica para ambos modelos ajustados con fines comparativos.

En la Tabla 3.2 se muestra la estimación en línea y suavizada para el valor esperado de la media de goles por partido (μ_t/n_t) y su varianza. Sólo se presentan resultados correspondientes a los años en los cuales comienza una década para efectos comparativos y se puede observar la evolución dinámica a lo largo del tiempo.

Figura 3.3: Promedio de Coles y Tendencia (en línea)

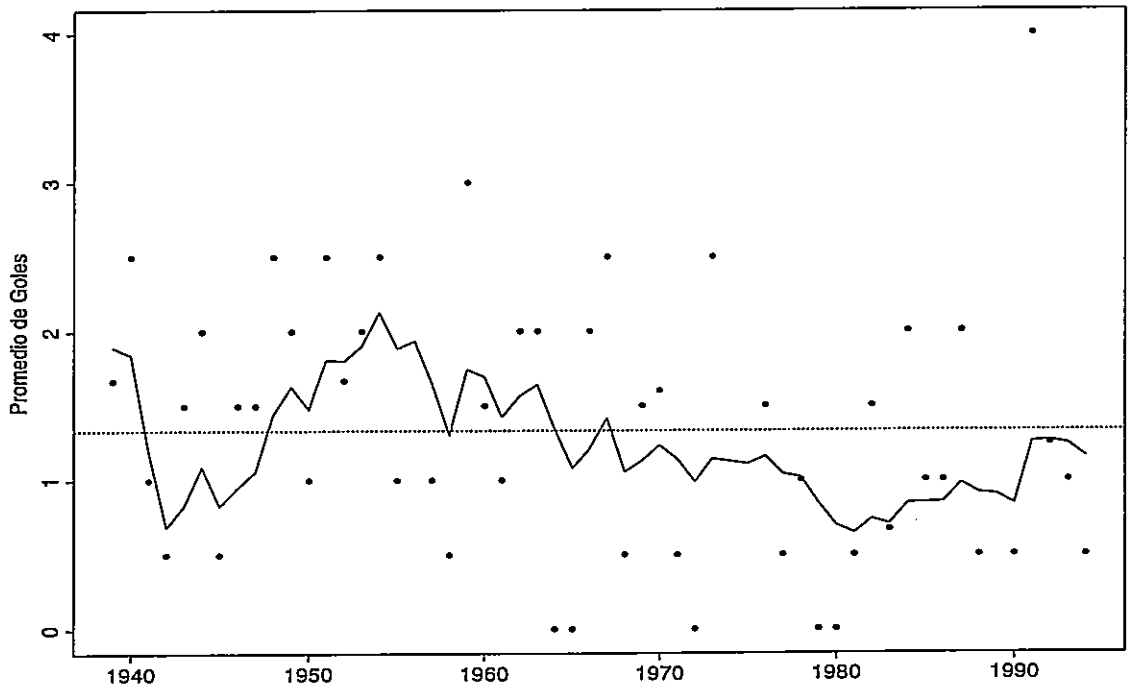
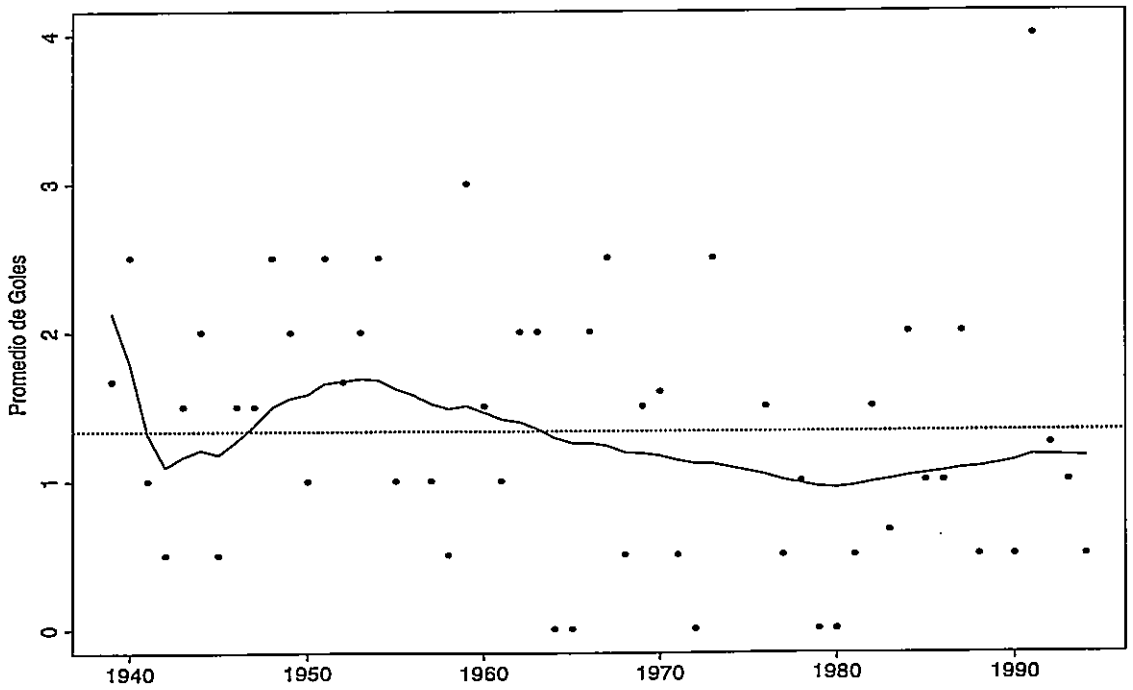


Figura 3.4: Promedio de Coles y Tendencia (suavizada)



3.2.4 Predicción

Suponga que para algún año t, en el algoritmo presentado en la Sección 3.2.2 en donde se describe la estimación de las componentes del modelo, se tiene que condicional en la información al tiempo $(t-1)$,

$$Y_t|\mu_t \sim Poisson(\mu_t), \qquad \mu_t|D_{t-1} \sim Gama(r_t, s_t) \tag{3.17}$$

entonces, a partir de los resultados mostrados en la Sección 2.4 o en la Sección 1.2.3, la distribución de probabilidades predictiva $Y_t|D_{t-1}$, está dado por

$$p(Y_t|D_{t-1}) = \int_0^\infty p(y_t|\mu_t)p(\mu_t|D_{t-1})d\mu_t = \frac{\Gamma(r_t+y_t)}{\Gamma(r_t)\Gamma(y_t+1)}p_t^{r_t}(1-p_t)^{y_t} \tag{3.18}$$

donde $y_t = 0, 1, 2, \ldots$; y

$$p_t = \frac{s_t}{s_t+1}.$$

Así, en cualquier año $(t-1)$ en el algoritmo del filtro de Kalman el sistema (3.12) entrega los

Tabla 3.3: **Probabilidades del Próximo Partido**

Número de Goles	en 1954	en 1981	en 1994	Con Promedio
0	0.13	0.55	0.33	0.26
1	0.24	0.32	0.35	0.35
2	0.25	0.10	0.20	0.23
3	0.18	0.02	0.08	0.10
4	0.11	0.00	0.03	0.03
5	0.05	0.00	0.01	0.01

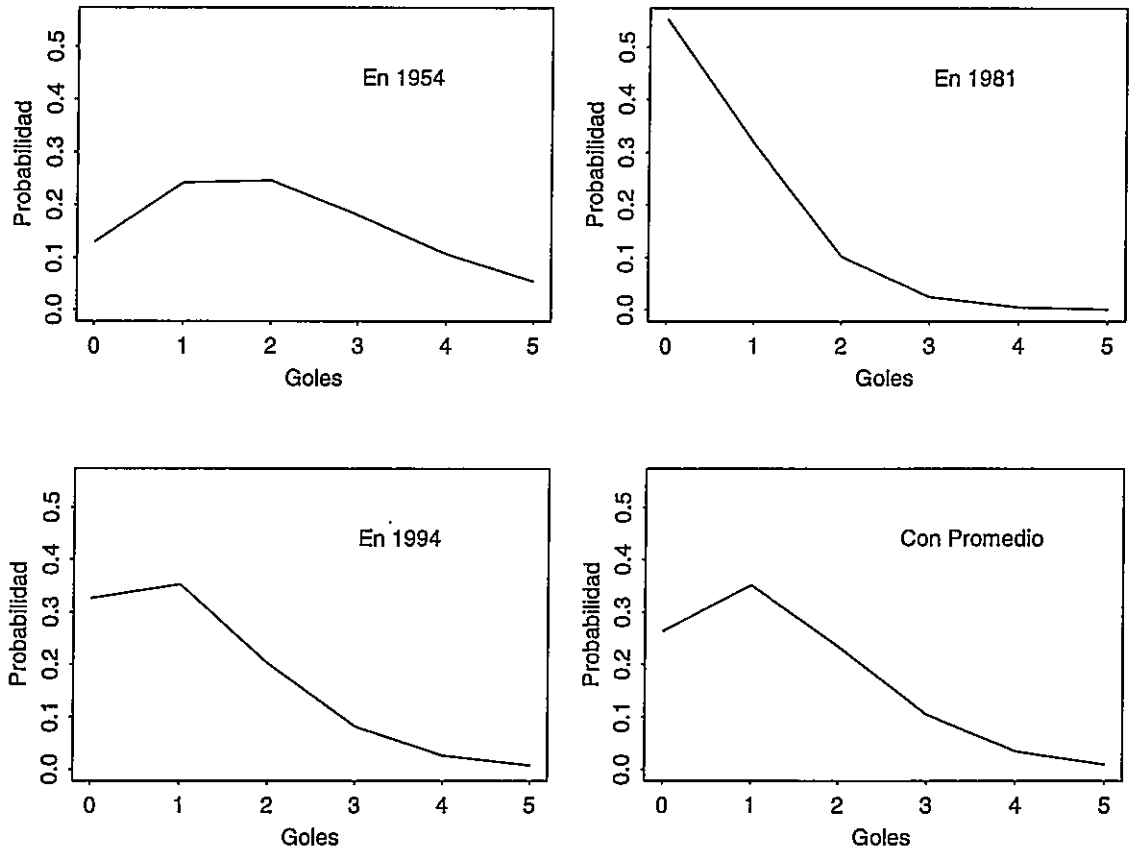
valores de r_t y s_t. Estos valores son todo lo necesario para calcular la distribución de probabilidades del número de goles del próximo año o en el próximo partido si se supone que la variable determinística n_t toma el valor **1**. La Tabla 3.3 muestra las probabilidades asociadas al núnero de goles en el próximo partido en los años 1954, 1981 y 1994, particularmente los dos primeros años estan determinados por el comportamiento del equipo en la cual tuvieron un rendimiento muy alto y bajo respectivamente. Una ilustración gráfica de estas probabilidades se muestra en la Figura 3.5

Desde luego que también se puede obtener probabilidades de predicción con horizontes de tiempo superiores a un año y/o para un número de partidos más que uno.

3.2.5 Residuo de Pearson

Para los modelos lineales generalizados no hay una definición clara para los residuos, mucho menos para el caso dinámico, porque, lo que se modela es el valor esperado de las observaciones o una transformación de ésta. Sin embargo, existen algunas definiciones que no gozan de consenso

Figura 3.5: **Probabilidad del Próximo Partido**



en la actualidad, uno de ellos al parecer el más adecuado para nuestro modelo, es el Residuo de Pearson, definido como en el Capítulo 2, dado por

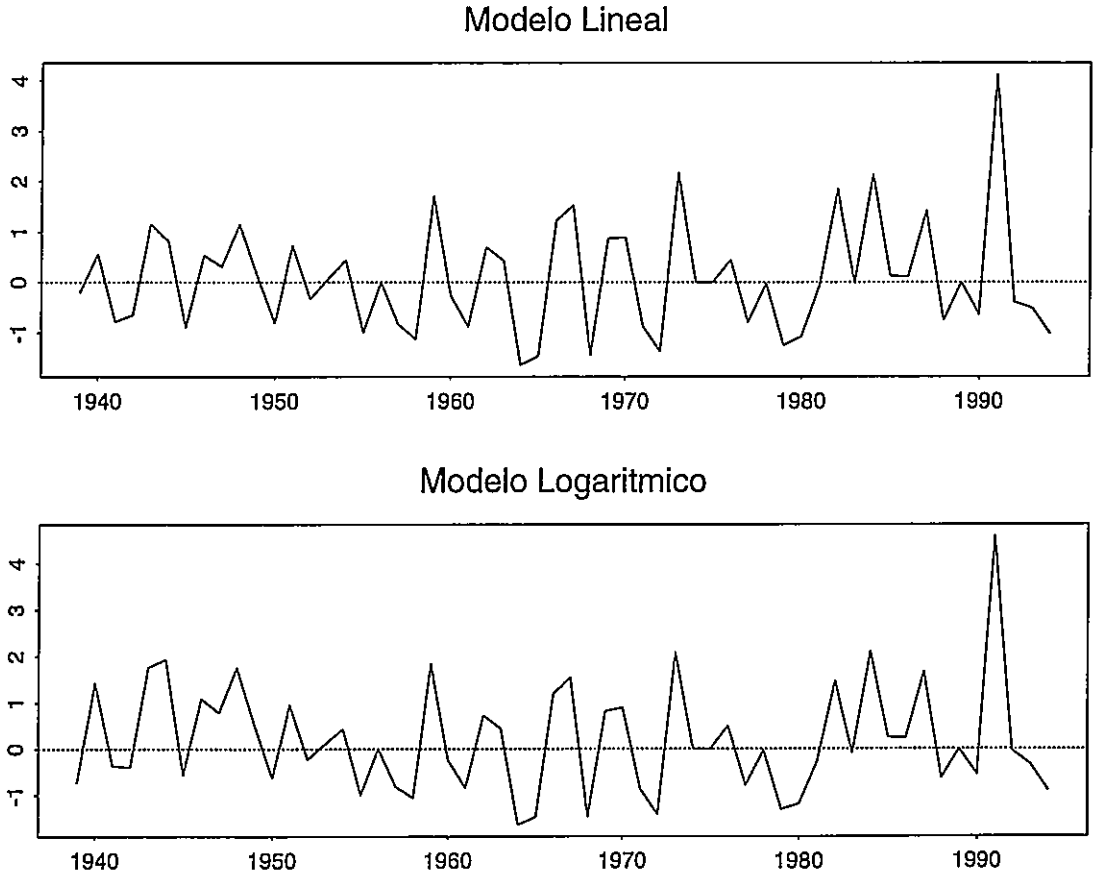
$$rp_t = \frac{y_t - \mathbb{E}(y_t|D_{t-1})}{\sqrt{\mathbb{V}(y_t|D_{t-1})}} \tag{3.19}$$

donde

$$\begin{aligned} \mathbb{E}(y_t|D_{t-1}) &= r_t(1-p_t)p_t^{-1} = \frac{r_t}{s_t} \\ \mathbb{V}(y_t|D_{t-1}) &= r_t(1-p_t)p_t^{-2} = \frac{r_t}{s_t^2}(s_t+1). \end{aligned}$$

Se define cero para el caso de los datos faltantes, y en este caso $\hat{y}_t$ se toma como el valor esperado de $\hat{\mu}/n_t$, con información al tiempo $t-1$, ó $t-2$ ó $\cdots$. La gráfica, de los residuos estimados para ambos modelos se muestran en la Figura 3.6.

Figura 3.6: **Residuos de Predicción o de Pearson para ambos modelos**



3.2.6 Función de Verosimilitud

La función de verosimilitud para los factores de descuento podemos escribir siguiendo la idea de la descomposición del error de predicción, ocupando para esto la densidad predictiva $p(y_t|D_{t-1})$ deducida en la Ecuación (3.18). Así, luego la función log-verosimilitud para δ_1 y δ_2 resulta

$$\ell(\delta_1, \delta_2|D_T) = \sum_{t=1}^{T} \log p(y_t|D_{t-1}). \tag{3.20}$$

Los resultados de las "devianzas" para ambos modelos, para alguna de las combinaciones de los parámetros se muestran en la Tabla 3.4, lo cual nos permite seleccionar un par de los δ_i óptimo. Sin embargo, a esto hay que agregar los resultados asociados con sus correspondientes residuos, para estimar finalmente los parámetros. En este estudio, de acuerdo a las pautas mencionadas, los factores de descuento óptimos resultaron $\delta_1 = 0.97$ y $\delta_2 = 0.95$ para ambos modelos, pero observando los residuos, nos quedamos con aquel modelo que usa la función de enlace natural en la proposición del modelo mostrado en la Ecuación (2.7).

Tabla 3.4: **Algunas devianzas para ambos modelos, basados en** $-2\log L$

Modelo Lineal			Modelo Logarítmico		
δ_1	δ_2	$-2\log L$	δ_1	δ_2	$-2\log L$
0.97	0.90	242.5576	0.97	0.90	243.8848
0.99	0.90	242.0256	0.99	0.90	243.4920
0.97	0.95	238.5639	0.97	0.95	241.3455

3.3 Conclusiones

El número de goles marcados por el equipo de fútbol de la Universidad Católica en los clásicos universitarios jugados contra el equipo de la Universidad de Chile es presentado en este trabajo como un proceso en el cual las observaciones tienen distribución Poisson, mientras que el logaritmo de la media evoluciona de acuerdo con un modelo de series de tiempo.

Las estimaciones muestran que un índice de la pericia del equipo de la U. Católica tiene una tendencia decreciente desde aproximadamente 1954 y hasta 1981. A partir de entonces el indicador mencionado crece pero en 1994 todavía no alcanzaba los niveles de 1954.

Las predicciones de las probabilidades del número de goles para el próximo partido son las siguientes. La probabilidad de no marcar goles es cercana a 0.30 y la probabilidad de marcar 1 gol tanbién es cercana a 0.30. La probabilidad de marcar dos goles es aproximadamente 0.2. Finalmente la probabilidad de marcar más de dos goles es aproximadamente 0.2.

Los resultados en cuanto a los residuos de predicción muestran que el modelo propuesto se ajusta razonablemente bien a los datos. La metodología propuesta abre la posibilidad de modelar series de tiempo para observaciones Poisson o con algunas extensiones menores de observaciones de otras distribuciones pertenecientes a la familia exponencial.

Tabla 3.5: Goles de los Clásicos Universitarios

Año	CDUC	U. de Chile	Año	CDUC	U. de Chile	Año	CDUC	U. de Chile
1939	0	2	1959	3	4	1975	-	-
	3	1	1960	2	4	1976	3	4
	2	0		1	2		0	0
1940	2	3	1961	0	0	1977	0	2
	3	2		0	0		1	1
1941	2	1		1	1	1978	1	1
	0	3		3	2		1	1
1942	0	2	1962	2	0	1979	0	0
	1	1		1	4		0	0
1943	1	1		3	5	1980	0	2
	2	1	1963	3	3		0	1
1944	2	4		1	1	1981	1	2
	2	2	1964	0	1		0	1
1945	1	4		0	0	1982	2	3
	0	2	1965	0	0		1	1
1946	2	2		0	0	1983	1	1
	1	0	1966	1	1		0	2
1947	3	2		3	5		1	1
	0	4	1967	3	2	1984	1	3
1948	3	3		2	3		3	2
	2	1	1968	0	2	1985	0	0
1949	4	1		0	1		2	1
	0	1		0	0	1986	1	1
1950	2	1		1	1		1	1
	0	1		1	0	1987	2	1
1951	3	1		1	2		2	1
	2	2	1969	3	4	1988	0	0
1952	0	1		0	1		1	0
	2	3		1	4	1989	-	-
	3	1		2	4	1990	0	0
1953	0	1	1970	4	1		1	1
	4	0		1	1	1991	5	2
1954	0	0		1	1		3	2
	5	0		0	1	1992	1	1
1955	1	3		2	1		0	0
	1	0	1971	0	2		1	3
1956	-	-		1	3		3	1
1957	1	1	1972	0	4	1993	1	1
	1	3		0	3		1	0
1958	1	3	1973	2	2	1994	1	0
	0	1		3	3		0	1
1959	3	3	1974	-	-			

Fuente: Club Deportivo Universidad Católica (CDUC)

Universidad Mayor de San Andres
Facultad de Ciencias Puras y Naturales
CARRERA DE MATEMATICA
BIBLIOTECA
LA PAZ - BOLIVIA

DIVISION BIENES E INVENTARIOS
VERIFICADO
Fecha
H. M. R
M. S

Capítulo 4

APLICACION II: RATING DE LA TELEVISION (TVR)

4.1 Análisis del Problema

En esta aplicación, como una nueva ilustración de la teoría desarrollada en el Capítulo 2, se estudia la evolución dinámica del rating de un programa de televisión (TVR). El *people meter* es un sistema mediante el cual se mide minuto a minuto, el rating de los programas de televisión en un hogar o en una persona. En una muestra de hogares seleccionados se coloca junto al televisor un dispositivo que transmite en cada minuto señales que permiten conocer los miembros del hogar que están frente al televisor y el programa de televisión sintonizado. Estas señales llegan a una oficina central que calcula, entre otras cosas, la proporción de hogares que ven cada uno de los programas que se está transmitiendo. Así el rating observado de un programa es una estimación de la proporción total de hogares que están sintonizando un programa.

Para analizar el comportamiento de los ratings, presentamos una metodología para el estudio de la evolución dinámica de los *ratings* de programas de televisión medidos minuto a minuto a través del sistema *people meter* implementado en la ciudad de Santiago de Chile. El objetivo principal está centrado en la estimación de las tendencias y en la predicción a corto plazo en los ratings.

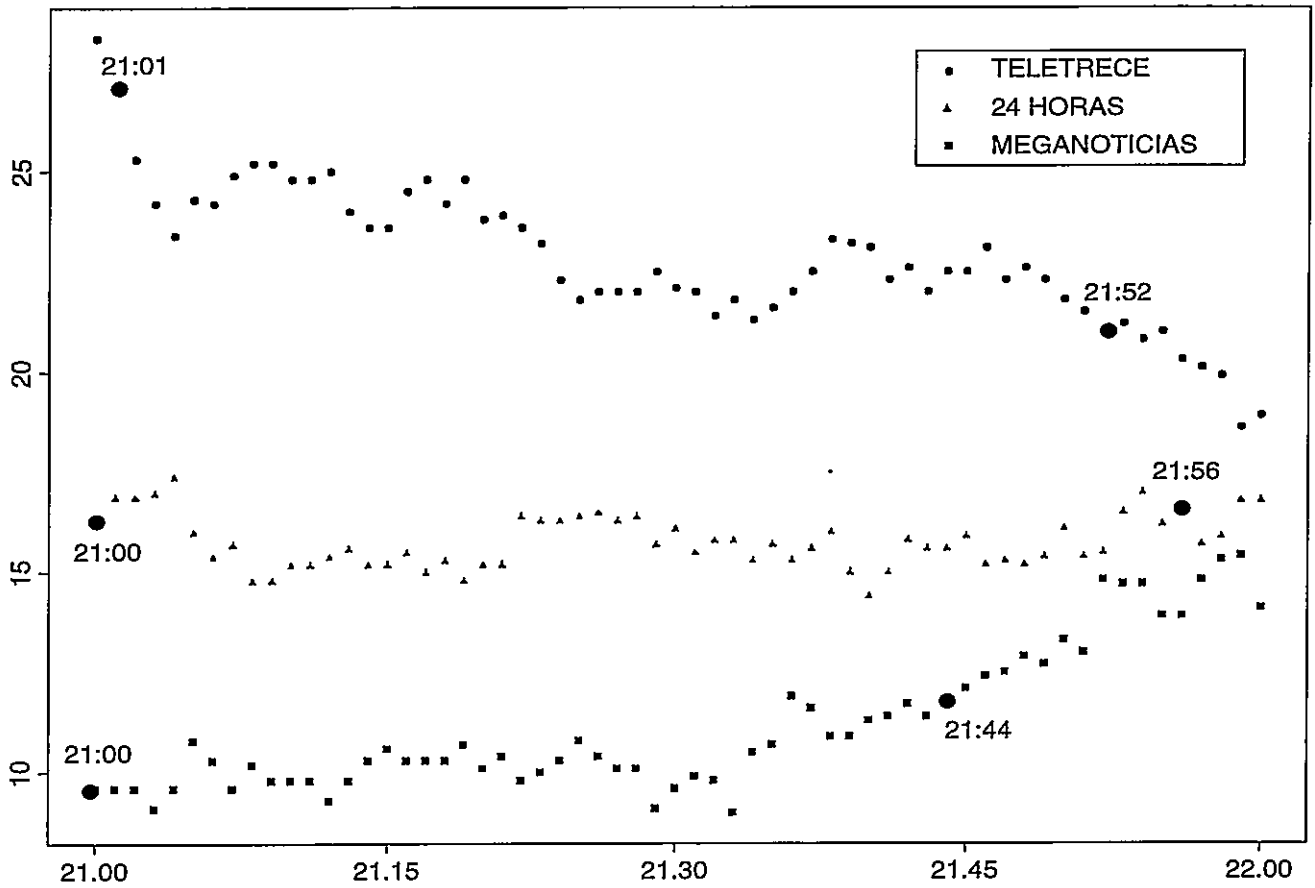
En este estudio se utilizan observaciones de los ratings de programas estelares de noticias entre las 21:00 y las 22:00 horas de tres canales de televisión y para un día que corresponde al 4 de Septiembre de 1995. Los tres programas que se estudian son:

- TELETRECE del Canal 13
- 24 HORAS del Canal 7
- MEGANOTICIAS del Canal 9

En la Figura 4.1 se presentan los ratings correspondientes a los tres programas de noticias del día Lunes mencionado. En esta representación se observan importantes fluctuaciones en los ratings

estimados minuto a minuto.

Figura 4.1: **Rating de Programas de Noticias**



Fuente: *Departamento de Estudios* de la Corporación de Televisión Universidad Católica de Chile, Canal 13, Santiago.

La naturaleza de las observaciones que se consideran en este estudio sugiere que en cada momento del tiempo el rating observado correspondiente a un programa puede ser considerado como una realización de una variable aleatoría con distribución Binomial de parámetros $n_t = 324$, que corresponde al tamaño de la muestra en el sistema *people meter* y μ_t que corresponde a la probabilidad de que en un hogar se esté viendo en el minuto t el programa. Es decir, μ_t corresponde al rating verdadero.

Ahora, la dependencia de las observaciones a través del tiempo aconseja el uso de una clase de modelos de series de tiempo para estudiar y predecir el comportamiento futuro de los ratings.

El rating observado en minutos consecutivos estan correlacionados y no pueden ser tratados como realizaciones independientes. En la literatura, no existen muchos estudios en los cuales se consideren series de tiempo con observaciones provenientes de la distribución Binomial. Desde un punto de vista metodológico, los modelos de series de tiempo que han predominado en la literatura estadística de los últimos años suponen distribución de probabilidades Normal para las observaciones, o restringen el análisis a los dos primeros momentos de la variables aleatorias, y por tanto, a estimaciones y predicciones lineales. En este caso, la aproximación Normal usual para la variable Binomial no resultó apropiada para describir el comportamiento de los ratings en varios ejercicios que se efectuaron.

La metodología propuesta en este estudio sigue de cerca los resultados en Harvey(1989), West y Harrison(1989) y también en West, Harrison y Migon(1985).

4.2 Un Modelo Dinámico Binomial–Beta

4.2.1 TVR de los Noticieros

Sea y_t el número de hogares que al tiempo t están viendo un determinado programa de televisión de un total de n_t hogares en la muestra. El rating observado del programa al tiempo t corresponde a la proporción

$$Rating_t = \frac{y_t}{n_t}, \qquad t = 1, 2, \ldots, T.$$

Supongamos que y_t es la realización de una variable aleatoria Binomial Y_t de parámetros n_t y μ_t, donde μ_t representa a la proporción esperada de los hogares que están viendo el programa. La Figura 4.1 y el conocimiento del proceso que generó las variables Y_t suguiere que μ_t deba ser considerada aleatoria con un cierto grado de evolución y cambio a través del tiempo. Como los cambios para los ratings no son predecibles, será mejor considerar a μ_t como una variable aleatoria. Para simplificar el análisis y los cálculos, supondremos que el parámetro μ_t tiene una distribución de probabilidades *Beta* de parámetros r_t y s_t. Así tenemos

$$Y_t|\mu_t \sim Bin(n_t, \mu_t), \qquad \mu_t|D_{t-1} \sim Beta(r_t, s_t) \tag{4.1}$$

donde D_{t-1} como se definió en los capítulos anteriores, contiene toda la información histórica. La función de densidad para la probabilidad μ_t tiene la forma

$$f(\mu_t) = \frac{\Gamma(r_t + s_t)}{\Gamma(r_t)\Gamma(s_t)} \mu_t^{r_t - 1}(1 - \mu_t)^{s_t - 1}$$

y es bien conocido que la media y la varianza para ésta distribución son

$$\mathbb{E}(\mu|D_{t-1}) = \frac{r_t}{r_t + s_t}, \qquad \mathbb{V}(\mu|D_{t-1}) = \frac{r_t s_t}{(r_t + s_t)^2(r_t + s_t + 1)}.$$

Este modelo conocido como Binomial–Beta requiere que en cada minuto t se especifiquen los valores de los parámetros r_t y s_t, los mismos que determinan completamente la distribución de

probabilidades para μ_t, que puede ser actualizado al tiempo t con el conocimiento de la obsevación $Y_t = y_t$ por los procedimientos descritos en la Sección de Ejemplos del Capítulo 2.

Como μ_t sólo está definido en el intervalo (0,1), consideramos las siguientes transformaciones

$$\lambda_t = g(\mu_t) = \begin{cases} \mu_t \\ \text{logit}(\mu_t) \end{cases} \tag{4.2}$$

donde

$$\text{logit}(\mu_t) = \log(\frac{\mu_t}{1-\mu_t}).$$

La primera especificación es la proporción esperada de hogares que están viendo el programa, mientras que la segunda especificación es el logaritmo de la razón entre la proporción esperada de los hogares que están viendo el programa y los que no están viendo el mismo programa. Así, la idea es proponer un modelo de series de tiempo, en la clase de modelos estructurales para la componente λ_t.

En la Figura 4.1 se observa que los ratings, y por tanto, una función del parámetro μ_t evoluciona linealmente en períodos cortos de tiempo. Así, para la componente λ_t se propone un modelo estructural llamado *local lineal* y que tiene la forma

$$\begin{aligned} \lambda_t &= \mu_t^* \\ \mu_t^* &= \mu_{t-1}^* + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t} \end{aligned} \tag{4.3}$$

Donde μ_t^* representa la media o la tendencia de λ_t. En otras palabras, esto puede ser interpretado como el nivel de la tendencia en la escala original o logito, según sea la función de enlace escogida. Esta componente evoluciona en función del minuto anterior más una componente β_{t-1} y en función de una componente aleatoria no predecible ω_{1t}. Donde la componente β_t que representa a la pendiente dinámica de la tendencia que evoluciona de acuerdo a un camino aleatorio, por cuanto ω_{2t} como ω_{1t} son ruido blancos de media cero y varianzas constantes y no correlacionado en el tiempo.

La especificación del modelo es completamente análoga a la utilizada en modelos para observaciones de la distribución Normal; ver Harvey(1989) y West y Harrison(1989). La diferencia aquí es que el modelo no está planteado para una observación con distribución Normal sino para una función del parámetro μ_t de la distribución Binomial.

4.2.2 Estimación

El modelo propuesto en la sección anterior es estimado para cada una de las series de rating. El algoritmo de estimación sigue las líneas descritas en el Capítulo 2. En este sentido se deben distinguir dos tipos de estimaciones. Cuando el programa de televisión está siendo transmitido, la información sobre los ratings llega minuto a minuto. La estimación del coeficiente μ_t se hace

con la información disponible hasta el minuto t, esto se conoce como la estimación *en línea*. Esta estimación da origen a predicciones del comportamiento del rating μ_t en los próximos minutos. Cuando llega la información correspondiente al minuto $(t+1)$, se actualizan las estimaciones y predicciones. En esta contexto, también puede tener sentido lo que se conoce como la estimación *suavizada*, el mismo que corresponde a la estimación de los rating cuando el programa haya finalizado su emisión al minuto T.

En el proceso de estimación se supone, en primer lugar, que las varianzas de las componentes ω_{1t} y ω_{2t} son conocidas, y constantes en el tiempo. Denotemos estas varianzas por σ_1^2 y σ_2^2 funcionalmente independientes asociados a cada componente μ_t^* y β_t respectivamente, y que los mismos pueden ser estimados usando la idea estándar de máxima verosimilitud cuando son desconocidas. Por otra parte, los datos usados para ajustar los modelos corresponden a los ratings entre el inicio y hasta su finalización del programa, excepto en el caso del noticiero TELETRECE del cuál tomamos los datos pasados 4 minutos, y para 24 HORAS pasado un minuto, es decir, desde 21:05 y 21:00 respectivamente, ya que en estos casos en los primeros minutos los datos estan fuertemente influenciados por un programa anterior.

El algoritmo que permite hacer las estimaciones y los cálculos para las componentes μ_t^* y β_t corresponde a una modificación de lo que se conoce como el filtro de Kalman, descrita anteriormente en el Capítulo 2. Para lo cual, primero escribimos el modelo propuesto en su forma de estado

$$\begin{aligned} \lambda_t &= F'\theta_t \\ \theta_t &= G\theta_{t-1} + \omega_t \end{aligned} \tag{4.4}$$

donde $\omega_t \sim [0, W_t]$, siendo $W_t =$ diag (σ_1^2, σ_2^2), y

$$F' = (1,0), \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \omega_t = \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix}, \qquad \theta_t = \begin{pmatrix} \mu_t^* \\ \beta_t \end{pmatrix}. \tag{4.5}$$

Filtro de Kalman

Información Inicial o a Posteriori Inicial

Los valores de m_0 y C_0, que usamos para inicializar el algoritmo para los tres programas en la escala logito, fueron como se muestra en la siguiente tabla, en donde los primeros valores de m_0 están escogidas de manera que sean coherentes en la escala de la función de enlace y para que el algoritmo sea estable desde el principio, una idea puede ser utilizar la información equivalente al tiempo $t = 0$ del día anterior o del mismo día de la semana anterior. Sin embargo, teóricamente podríamos dar inclusive un valor inicial con distribución difusa, es decir, la media m_0 cualquiera y la varianza arbitrariamente grande, pero esto provocaría en algunos casos inestabilidades numéricas computacionales.

Momentos Iniciales	TELETRECE Canal 13	24 HORAS Canal 7	MEGANOTICIAS Canal 9
m_0	$\begin{pmatrix} -1.11 \\ 0.00 \end{pmatrix}$	$\begin{pmatrix} -1.59 \\ 0.00 \end{pmatrix}$	$\begin{pmatrix} -2.25 \\ 0.00 \end{pmatrix}$
C_0	$\begin{pmatrix} 0.1 & 0.0 \\ 0.0 & 0.1 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0.00 \\ 0.00 & 0.01 \end{pmatrix}$	$\begin{pmatrix} 0.01 & 0.00 \\ 0.00 & 0.01 \end{pmatrix}$

Así

$$(\theta_0|D_0) \sim [m_0, C_0]. \tag{4.6}$$

Paso 1: A Priori para λ_1
Ahora, de la ecuación de transición (4.4), se sigue que los momentos de la distribución a priori de θ_1 condicional en la información histórica son:

$$\begin{aligned} a_1 &= Gm_1 \\ R_1 &= GC_0G' + W_1 \end{aligned} \tag{4.7}$$

de modo que $(\theta_1|D_0) \sim [a_1, R_1]$ y $W_1 =$ diag $(\hat{\sigma}_1^2, \hat{\sigma}_2^2)$ diag (C_0). Donde las varianzas de los errores mostradas en la Tabla 4.3 fueron estimados siguiendo la idea de máxima verosimilitud, de modo que los residuos estimados sean efectivamente un ruido blanco.

Luego, siempre condicional en la información disponible al tiempo $t = 0$, ocupando la ecuación de medida (4.4) se puede obtener la media y la varianza de λ_1. Esto es

$$\begin{aligned} f_1 &= \mathbb{E}(\lambda_1|D_0) = F'a_1 \\ q_1 &= \mathbb{V}(\lambda_1|D_0) = F'R_1F. \end{aligned} \tag{4.8}$$

Paso 2: Especificación de la Distribución a Priori
En este caso, los parámetros de la distribución a priori conjugada para el caso del la función de enlace indentidad satisfacen el sistema

$$\mathbb{E}(\mu_1|D_0) = \frac{r_1}{r_1 + s_1} = f_1, \qquad \mathbb{V}(\mu_1|D_0) = \frac{f_1(1 - f_1)}{r_1 + s_1 + 1} = q_1$$

siendo n_1 conocido, estas ecuaciones definen de manera única los parámetros r_1 y s_1 y por lo tanto la distribución de probabilidades para μ_1 siempre condicional en la información disponible al tiempo $t = 0$.

Por otra parte, si la función de enlace entre λ_t y μ_t es el logito, el sistema de ecuaciones que satisfacen los parámetros r_1 y s_1, de acuerdo con el Ejemplo 2.5.2 será

$$\begin{aligned} \gamma(r_1) - \gamma(s_1) &= f_1 \\ \dot{\gamma}(r_1) + \dot{\gamma}(s_1) &= q_1. \end{aligned} \tag{4.9}$$

Como en el caso anterior, estas ecuaciones determinan de manera única los valores de r_1 y s_1 y por tanto la distribución de probabilidades de μ_1 condicional en la información contenida en D_0.

Sin embargo, la resolución del sistema de ecuaciones (4.9) no es inmediato, pero afortunadamente en este caso, como los valores de los parámetros no serán pequeños, por lo que podemos usar una aproximación logarítmica para la función digamma $\gamma(\cdot)$, luego las soluciones, ver el Ejemplo 2.5.2, se pueden escribir de la siguiente manera

$$r_1 = \frac{1+\exp(f_1)}{q_1}, \qquad s_1 = \frac{1+\exp(-f_1)}{q_1}.$$

Paso 3: La Distribución a Posteriori para μ_1
Al tiempo $t = 1$ conocemos la observación y_1, o la proporción de hogares que están viendo el programa. De manera análoga como en la Ecuación (2.4), utilizando el Teorema de Bayes, se demuestra que la distribución a posteriori de μ_1, condicional en la observación y_1, es

$$\mu_1|D_1 \sim Beta[r_1 + y_1, s_1 + n_1 - y_1]. \tag{4.10}$$

Esta distribución de probabilidades muestra un grado de incertidumbre respecto de la proproción μ_1. La observación y_1 afecta a esta distribución en el sentido de que la misma es formada utilizando la información previa al tiempo $t = 0$ y la observación o la proporción al tiempo $t = 1$.

Ahora, la relación $\lambda_t = g(\mu_t)$, especificada en el modelo nos permite encontrar los momentos posteriores para λ_1. Así, para el caso del enlace *identidad* resulta

$$f_1^* = \mathbb{E}(\lambda_1|D_1) = \frac{r_1 + y_1}{r_1 + s_1 + n_1}, \qquad q_1^* = \mathbb{V}(\lambda_1|D_1) = \frac{f_1^*(1-f_t^*)}{r_1 + s_1 + n_1 + 1}$$

mientras que para el modelo con enlace natural logito, los momentos posteriores como se muestra en el Apéndice A estan dados por

$$\begin{aligned} f_1^* &= \mathbb{E}(\lambda_1|D_1) = \gamma(r_1 + y_1) - \gamma(s_1 + n_1 - y_1) \\ q_1^* &= \mathbb{V}(\lambda_1|D_1) = \dot{\gamma}(r_1 + y_1) + \dot{\gamma}(s_1 + n_1 - y_1). \end{aligned} \tag{4.11}$$

Paso 4: Momentos a Posteriori para θ_t
Para terminar el ciclo recursivo y disponer de la media y la varianza de las componentes del vector de estado θ_1 condicional en la información $t = 1$. Aplicando las relaciones deducidas en la Sección 2.3, resulta que $(\theta_1|D_1) \sim [m_1, C_1]$. Donde

$$\begin{aligned} m_1 &= \mathbb{E}(\theta_1|D_1) = a_1 + R_1 F(f_1^* - f_1)/q_1 \\ C_1 &= \mathbb{V}(\theta_1|D_1) = R_1 - R_1 F F' R_1 (1 - \frac{q_1^*}{q_1})/q_1. \end{aligned}$$

El procedimiento general propuesto para $t = 0$, se puede repetir en cada momento del tiempo según se obtengan nuevas observaciones. Es decir, los valores iniciales de entrada para un tiempo t son los momentos posteriores al tiempo $(t-1)$. Así se puede obtener en cada minuto una estimación de las componentes λ_t y de μ_t, como también de β_t.

4.2.3 Nivel de la Tendencia en Línea y Suavizada

La estimación en línea de la proporción esperada μ_t queda determinada por la media de la distribución a posteriori de la Ecuación (4.10) aplicada para cada tiempo t, cuya representación gráfica para cada programa en estudio se muestran en la Figura 4.2 para el programa TELETRECE, en la Figura 4.4 para el noticiero 24 HORAS, y en la Figura 4.6 para MEGANOTICIAS, en donde las proporciones estimadas están multiplicadas por 100, donde la media a posteriori para μ_t está dada por

$$\mathbb{E}(\mu_t|D_t) = \frac{r_t + y_t}{r_t + s_t + n_t} \tag{4.12}$$

Tabla 4.1: **Estimación de la Tendencia en Línea y Suavizada**

Programa	Hora	Proporción Rating Observado	Estimación en Línea $\mathbb{E}(\mu_t\|D_t)$	Estimación en Línea $\mathbb{V}(\mu_t\|D_t)$	Estimación Suavizada $\mathbb{E}(\mu_t\|D_T)$	Estimación Suavizada $\mathbb{V}(\mu_t\|D_T)$
TELETRECE	21:05	24.3	24.51	3.118	24.54	2.406
	21:10	24.8	25.03	3.542	24.94	2.542
	21:20	23.8	24.21	3.449	24.12	2.474
	21:30	22.1	22.17	3.200	22.06	2.283
	21:40	23.1	23.34	3.357	23.01	2.371
	21:50	21.8	22.01	3.188	21.91	2.299
	21:52	21.0	21.14	3.073	21.14	3.084
24 HORAS	21:05	16.0	16.65	2.154	16.33	1.377
	21:10	15.2	14.81	1.823	15.09	1.236
	21:20	15.2	15.08	1.877	15.30	1.261
	21:30	16.1	16.10	2.007	15.99	1.315
	21:40	14.4	14.88	1.862	15.13	1.257
	21:50	16.1	15.68	1.943	15.16	1.279
	21:56	16.6	16.66	2.079	16.66	2.087
MEGANOTICIAS	21:05	10.8	10.07	1.302	9.94	0.811
	21:10	9.8	9.94	1.241	9.90	0.798
	21:20	10.1	10.41	1.305	10.33	0.834
	21:30	9.6	9.56	1.184	9.77	0.791
	21:40	11.3	11.40	1.444	11.35	0.933
	21:44	11.8	11.73	1.471	11.73	1.479

Figura 4.2: **Estimación en Línea: Programa TELETRECE**

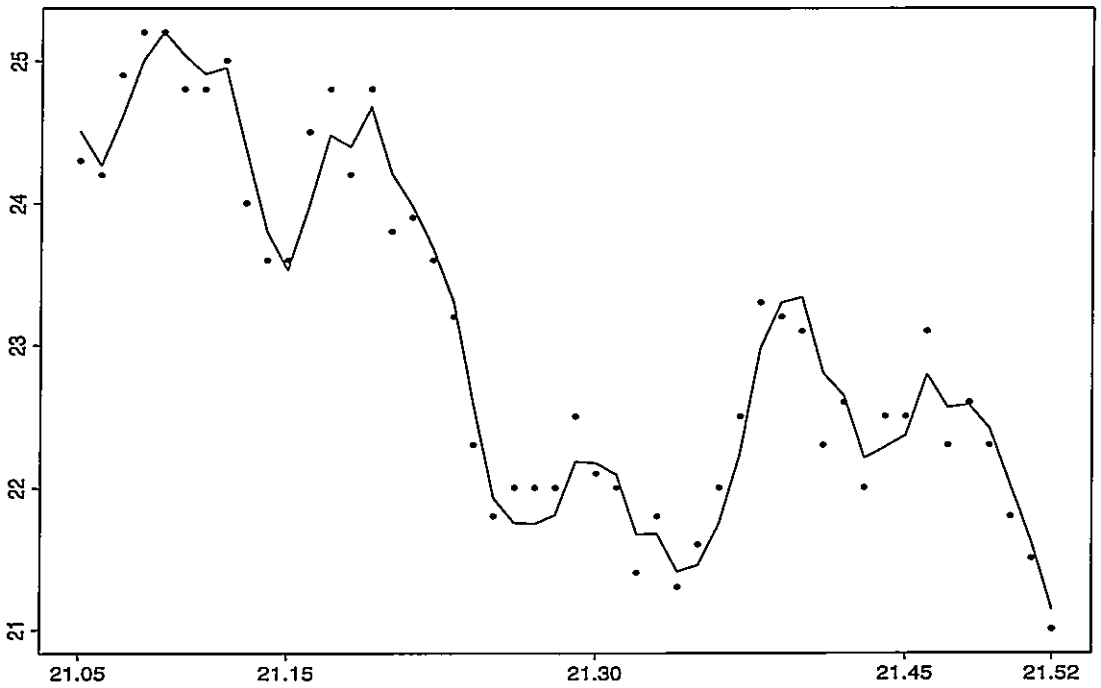


Figura 4.3: **Estimación Suavizada: Programa TELETRECE**

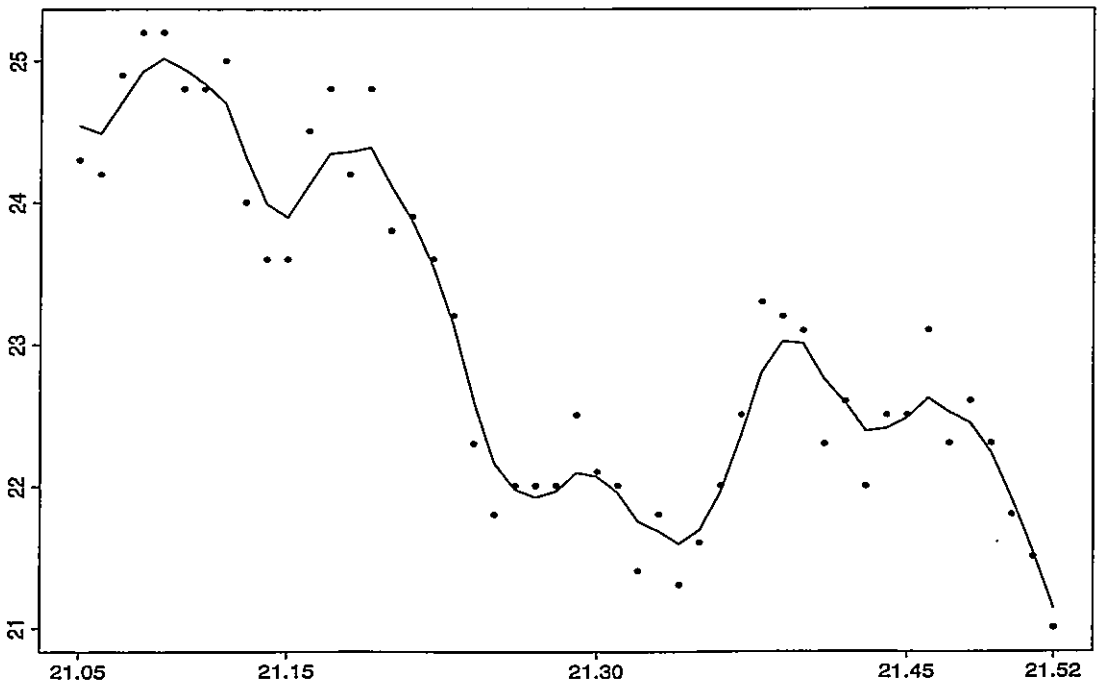


Figura 4.4: **Estimación en Línea: Programa 24 HORAS**

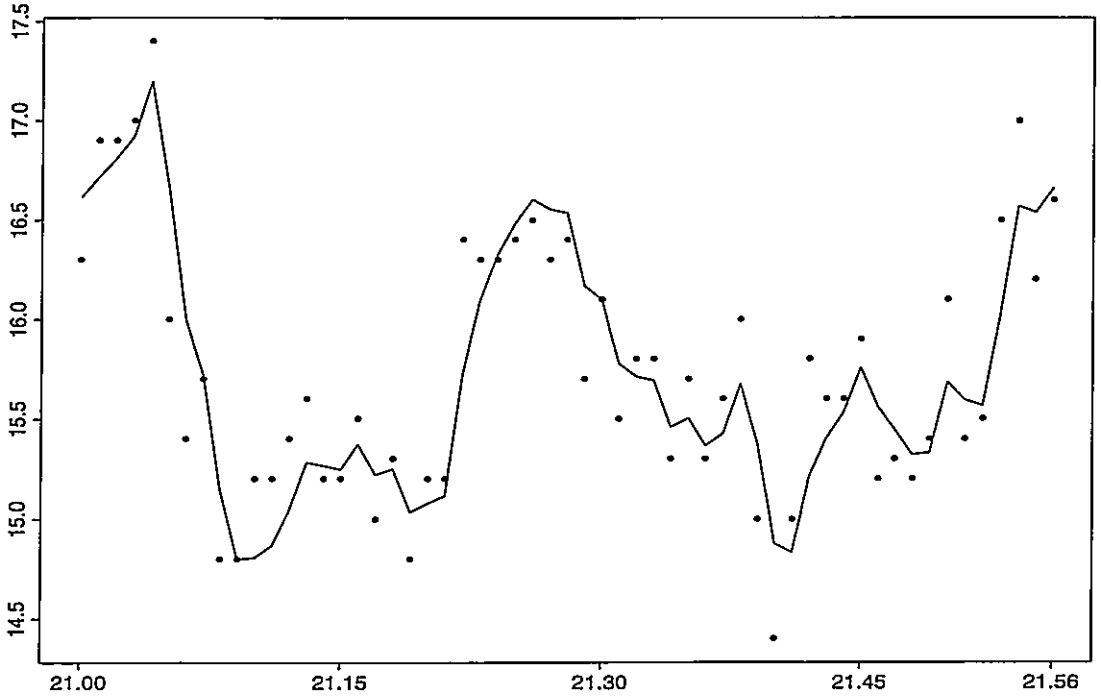


Figura 4.5: **Estimación Suavizada: Programa 24 HORAS**

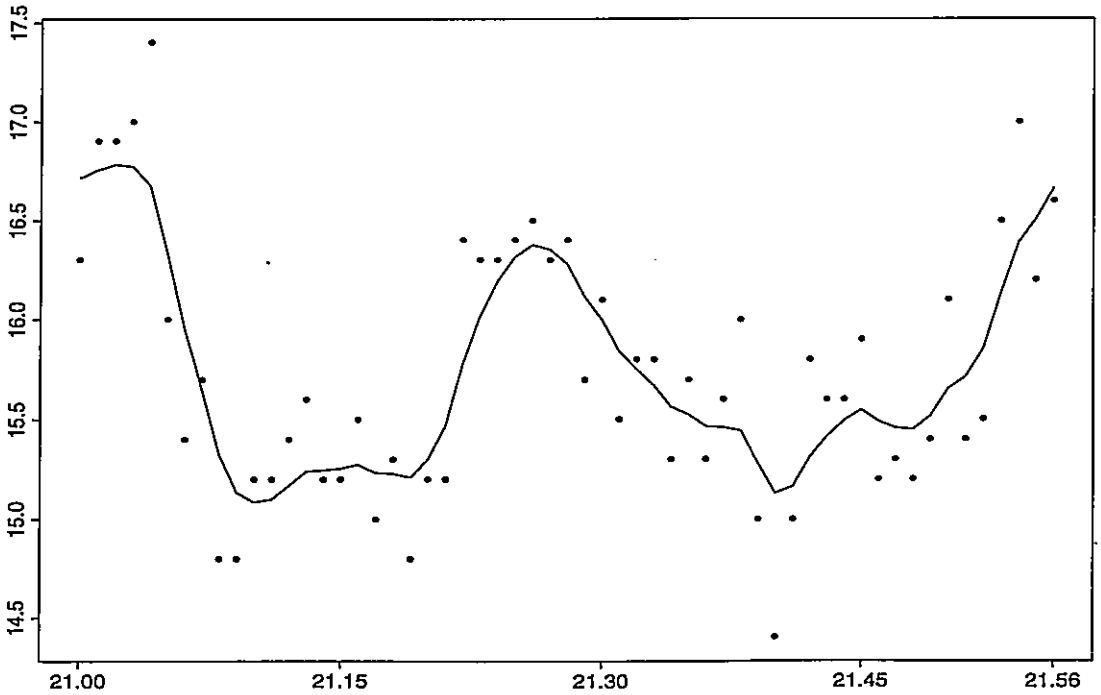


Figura 4.6: **Estimación en Línea: Programa** MEGANOTICIAS

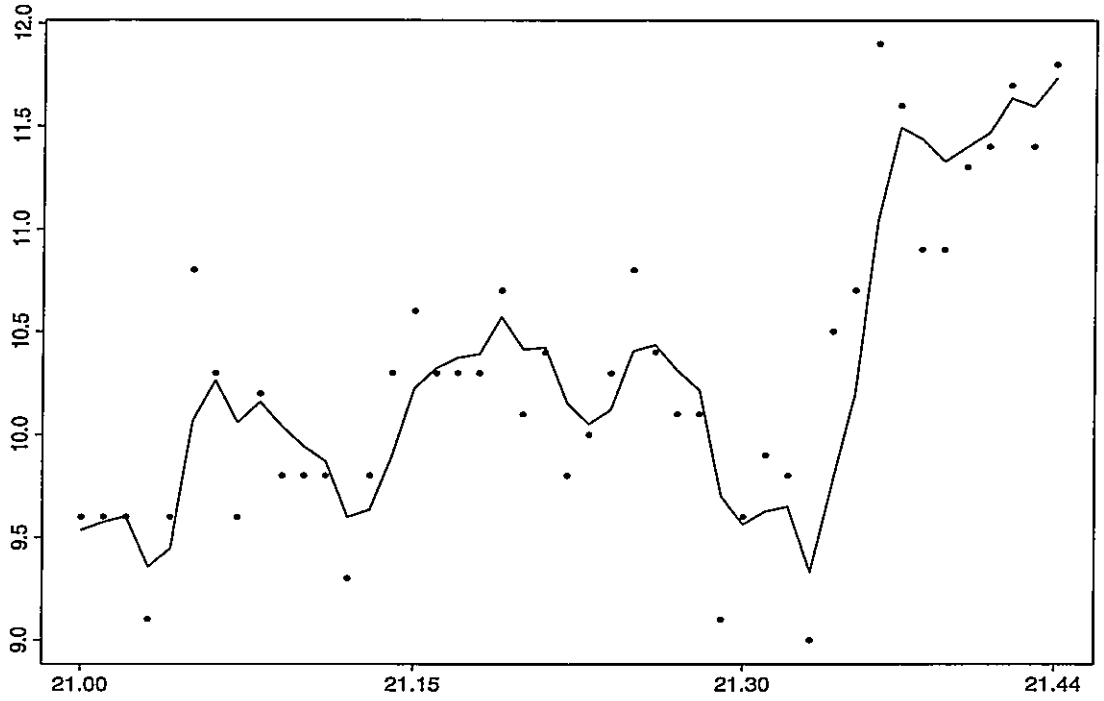
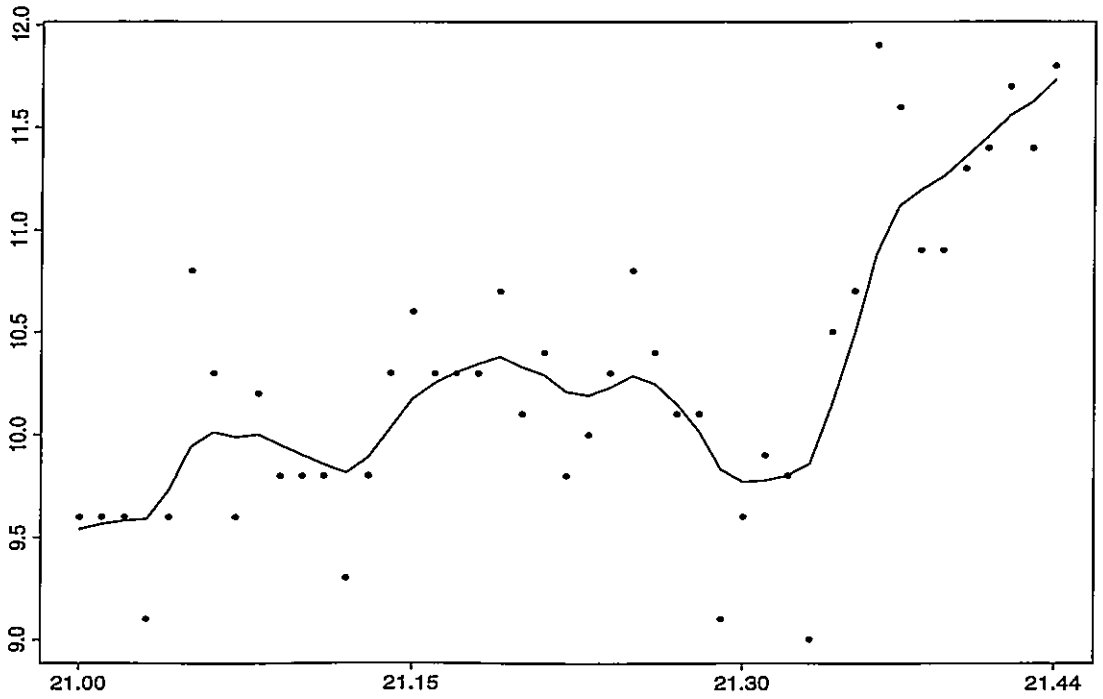


Figura 4.7: **Estimación Suavizada: Programa** MEGANOTICIAS



Un procedimiento similar se puede seguir para el caso del nivel suavizado usando toda la información disponible hasta el tiempo t

$$\mathbb{E}(\mu_{t-k}|D_t) = \frac{r_t(-k) + y_{t-k}}{r_{t-k} + s_t(-k) + n_{t-k}}, \qquad (1 \leq k < T). \tag{4.13}$$

Los resultados de las estimaciones del nivel suavizado de la tendencia, se muestran en la Figura 4.3 para el Canal 13, en la Figura 4.5 para el Canal 7, y en la Figura 4.7 para el Canal 9. De estos se desprende que la proporción de rating para el Canal 13 tiene una tendencia decreciente aunque hay momentos donde se observan repuntes importantes después de una cuarta hora de haberse iniciado el programa como también después de una media hora, y aún así se mantiene por encima de los otros programas de otros canales. Con respecto del programa de 24 HORAS la variación del nivel de rating estimado sólo registra aproximadamente un cambio de un 2%, presentando alzas a la mitad del programa como también a la finalización de la misma. Finalmente con respecto del noticiero MEGANOTICIAS del Canal 9, se observa aún menos variación en la primera media hora, en donde la variación en el rating fluctua aproximadamente solo en un 1%, registrandose al final del programa un crecimiento en un orden de mas o menos del 2%. En este trabajo sólo se presentan los resultados para el caso en que λ_t corresponde al logito de la proporción esperada.

En la Tabla 4.1 se muestran las estimaciónes en línea y suavizada para para el valor esperado del rating μ_t y sus respectivas varianzas, solo se muestran resultados correspondientes a cada 10 minutos para efectos comparativos entre las distintas programas y en ésta tabla se puede observar claramente la evolución dinámica de los rating a través del tiempo.

4.2.4 Predicción

Recordemos que uno de los objetivos del trabajo consiste en predecir el nivel de los rating a corto plazo. Para esto, suponga que para algún minuto t, en el algoritmo presentado en la Sección 4.2.2 donde se describe la estimación de las componentes del modelo, se tiene que condicional en la información al minuto anterior, $(t-1)$,

$$Y_t|\mu_t \sim Bin(n_t, \mu_t), \qquad \mu_t|D_{t-1} \sim Beta(r_t, s_t) \tag{4.14}$$

entonces, a partir de los resultados mostrados en la Sección 2.4 o en la Sección 1.2.3, la distribución de probabilidades predictiva $Y_t|D_{t-1}$, lo que nos permitirá además escribir la función de verosimilitud para los hiperparámetros, está dada por una distribución de probabilidades conocida como Beta-Binomial. Esto es,

$$\begin{aligned} p(Y_t|D_{t-1}) &= \int_0^1 p(y_t|\mu_t)p(\mu_t|D_{t-1})d\mu_t, \qquad y_t = 0, 1, 2, \ldots, n_t \\ &= \frac{\Gamma(r_t + s_t)}{\Gamma(r_t)\Gamma(s_t)\Gamma(r_t + s_t + n_t)} \binom{n_t}{y_t} \Gamma(r_t + y_t)\Gamma(s_t + n_t - y_t). \end{aligned} \tag{4.15}$$

Ahora en realidad, por la naturaleza del problema, lo que realmente interesa es predecir la proporción más que el número de hogares que ven un determinado programa. Es decir, el interés

está puesto en predecir μ_t. Por lo que debemos más bien considerar $(\mu_{t+k}|D_t)$, para $k = 1, 2 \ldots$. Con los procedimientos descritos en la Sección 2.4, una vez calculada los valores de $r_t(k)$ y $s_t(k)$, la predicción a k–pasos para la proporción esparada con información al minuto t será

$$\hat{\mu}_t(k) = \frac{r_t(k)}{r_t(k) + s_t(k)}, \qquad k = 1, 2, 3, \ldots, \tag{4.16}$$

Tabla 4.2: Predicciones de los Ratings

	Teletrece				24 Horas				Meganoticias			
k	Hora	$\hat{\mu}_t(k)$	p_t	$\hat{e}_t(k)$	Hora	$\hat{\mu}_t(k)$	p_t	$\hat{e}_t(k)$	Hora	$\hat{\mu}_t(k)$	p_t	$\hat{e}_t(k)$
0	21.13	24.37	24.0	-0.37190	21.07	15.72	15.7	-0.02045	21.10	9.943	9.8	-0.14343
1	21.14	24.12	23.6	-0.52175	21.08	15.46	14.8	-0.66080	21.11	9.931	9.8	-0.13052
2	21.15	23.87	23.6	-0.27335	21.09	15.20	14.8	-0.40466	21.12	9.918	9.3	-0.61763
3	21.16	23.63	24.5	0.87329	21.10	14.95	15.2	0.24798	21.13	9.905	9.8	-0.61763
4	21.17	23.38	24.8	1.41816	21.11	14.70	15.2	0.49716	21.14	9.892	10.3	0.40812
0	21.17	24.48	24.8	0.31983	21.17	15.22	15.0	-0.21989	21.17	10.376	10.3	-0.07590
1	21.18	24.70	24.2	-0.50201	21.18	15.20	15.3	0.09546	21.18	10.472	10.3	-0.17201
2	21.19	24.93	24.8	-0.12520	21.19	15.19	14.8	-0.38921	21.19	10.569	10.7	0.13110
3	21.20	25.15	23.8	-1.34974	21.20	15.17	15.2	0.02612	21.20	10.667	10.1	-0.56658
4	21.21	25.38	23.9	-1.47561	21.21	15.16	15.2	0.04143	21.21	10.765	10.4	-0.36506
0	21.23	23.31	23.2	-0.10706	21.33	15.69	15.8	0.10866	21.29	9.700	9.1	-0.60043
1	21.24	23.02	22.3	-0.72351	21.34	15.60	15.3	-0.30240	21.30	9.538	9.6	0.06159
2	21.25	22.74	21.8	-0.94240	21.35	15.51	15.7	0.18613	21.31	9.379	9.9	0.52119
3	21.26	22.46	22.0	-0.46371	21.36	15.43	15.3	-0.12575	21.32	9.222	9.8	0.57838
4	21.27	22.19	22.0	-0.18746	21.37	15.34	15.6	0.26196	21.33	9.067	9.0	-0.06679
0	21.32	21.67	21.4	-0.27308	21.44	15.53	15.6	0.06625	21.35	10.198	10.7	0.50240
1	21.33	21.49	21.8	0.30709	21.45	15.62	15.9	0.27715	21.36	10.343	11.9	1.55717
2	21.34	21.31	21.3	-0.01384	21.46	15.71	15.2	-0.51238	21.37	10.490	11.6	1.11012
3	21.35	21.14	21.6	0.46415	21.47	15.80	15.3	-0.50232	21.38	10.639	10.9	0.26122
4	21.36	20.96	22.0	1.04105	21.48	15.89	15.2	-0.69267	21.39	10.790	10.9	0.11046
0	21.37	22.23	22.5	0.26727	21.53	16.02	16.5	0.47971	21.41	11.469	11.4	-0.06895
1	21.38	22.49	23.3	0.81086	21.54	16.17	17.0	0.82579	21.42	11.582	11.7	0.11760
2	21.39	22.75	23.2	0.45236	21.55	16.33	16.2	-0.12933	21.43	11.697	11.4	-0.29681
3	21.40	23.01	23.1	0.09176	21.56	16.49	16.6	0.11437	21.44	11.812	11.8	-0.01221
4	21.41	23.27	22.3	-0.97092	21.57	16.64	15.7	-0.94000	21.45	11.929	12.1	0.17100

p_t: proporción de rating observada, $\hat{e}_t(k)$: error de predicción

mientras que la varianza asociada a esta predicción está dada por

$$\mathbb{V}[\mu_t(k)|D_t)] = \frac{r_t(k)s_t(k)}{(r_t(k) + s_t(k))^2(r_t(k) + s_t(k) + 1)}.$$

Haciendo uso de la fórmula anterior, la predicción en algunos minutos escogidos con horizontes de 4 minutos, durante el desarrollo de los programas, se muestran en la Tabla 4.2.

4.2.5 Residuo de Pearson

Como en la aplicación anterior el residuo de Pearson definido en el Capítulo 2, es

$$rp_t = \frac{y_t - \mathbb{E}(y_t|D_{t-1})}{\sqrt{\mathbb{V}(y_t|D_{t-1})}} \tag{4.17}$$

donde

$$\begin{aligned} \mathbb{E}(y_t|D_{t-1}) &= n_t \frac{r_t}{r_t + s_t} \\ \mathbb{V}(y_t|D_{t-1}) &= \frac{n_t r_t s_t}{(r_t + s_t)^2} \times \frac{r_t + s_t + n_t}{r_t + s_t + 1}. \end{aligned}$$

La gráfica, de los residuos estimados para los tres programas de televisión se muestran en las Figuras 4.8, 4.9 y 4.10.

4.2.6 Función de Verosimilitud

La función de verosimilitud para las varianzas de los errores σ_1^2 y σ_2^2 podemos escribir siguiendo la idea de la descomposición del error de predicción, ocupando para esto la densidad predictiva $p(y_t|D_{t-1})$ deducida en la Ecuación (4.15). Así, después de algunas simplificaciones, la función log–verosimilitud para σ_1^2 y σ_2^2 resulta

$$\ell(\sigma_1^2, \sigma_2^2|D_T) = \sum_{t=1}^{T} \log p(y_t|D_{t-1}). \tag{4.18}$$

Tabla 4.3: **Varianzas Estimadas**

Programa	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$
TELETRECE	10^{-10}	0.003
24 HORAS	10^{-10}	0.001
MEGANOTICIAS	10^{-10}	0.001

Después de haber ajustado los modelos con los datos, las estimaciones para las varianzas por máxima verosimilitud Ecuación (4.18) resultaron como se muestra en la Tabla 4.3 de modo que los residuos resulten un ruido blanco.

Figura 4.8: **Residuos de Pearson: Programa** TELETRECE

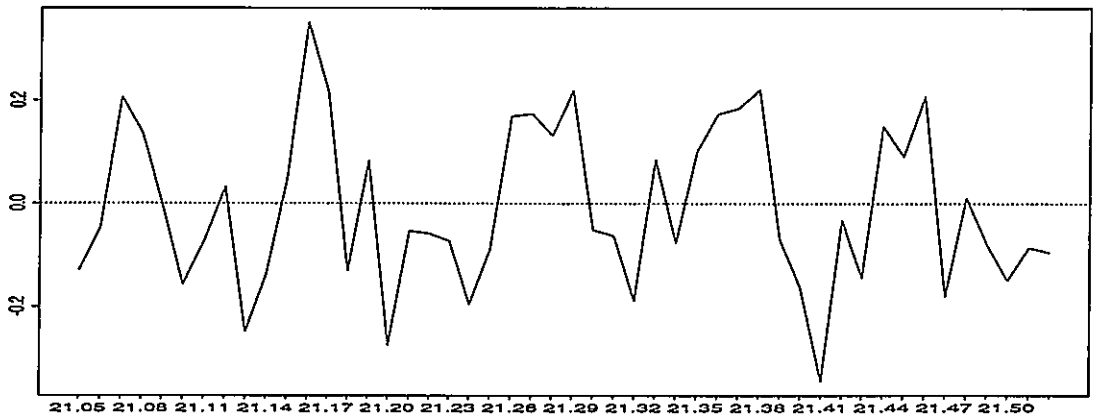


Figura 4.9: **Residuos de Pearson: Programa** 24 HORAS

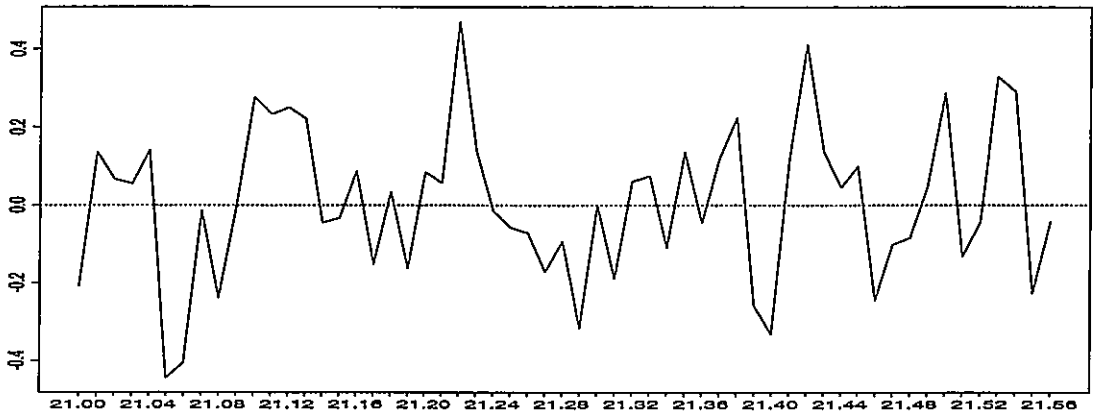
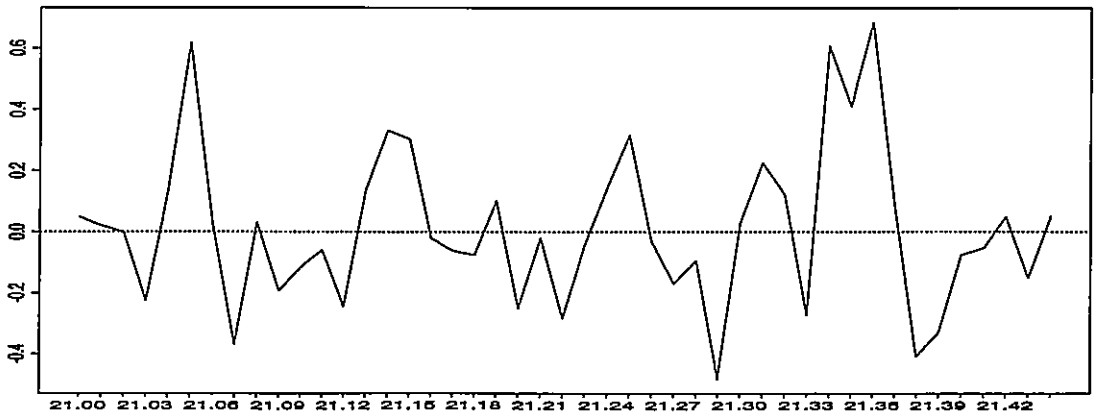


Figura 4.10: **Residuos de Pearson: Programa** MEGANOTICIAS



4.3 Conclusiones

En términos generales, los ratings, minuto a minuto, correspondientes a los tres programas de televisión son representados como un proceso en el cual las observaciones tienen distribución Binomial, mientras que una transformación de la media de esta distribución evoluciona de acuerdo con un modelo de series de tiempo.

Las estimaciones en línea y suavizada muestran que el modelo puede ser utilizado para proyectar tendencias a corto plazo. Se puede observar además que para ese día en particular, los ratings del Canal 13 se mantiene por encima de los otros dos noticieros en todo momento del desarrollo del programa, mientras que del Canal 7 se mantiene también por encima que del Canal 9. Por otra parte, la tendencia de los ratings del programa TELETRECE tiene en general un comportamiento decreciente, mientras que de MEGANOTICIAS ocurre lo contrario y de 24 HORAS se mantiene en un nivel sin muchas variaciones.

El comportamiento general de TELETRECE presenta un decaimiento en principio porque este viene influenciada por un programa anterior, que es, una teleserie muy atractiva, luego pasado unos 5 minutos después de haberse iniciado el programa, la tendencia presenta una recuperación por unos 10 minutos alcanzando un nivel de 25% en el rating hasta que en 21:35 horas llega a un nivel no menor a 21%, luego presenta nuevamente presenta una recuperación considerable por unos 5 minutos y termina con un nivel más bajo, aunque aún se mantiene en un nivel superior en comparación con otros programas competidoras del mismo tipo. Una manera de ver gráficamente el comportamiento es mirar la gráfica de la tendencia suavizada en la Figura 4.3.

El programa 24 HORAS, que es uno de los fuertes competidores del TELETRECE, presenta un buen inicio llegando la tendencia inclusive hasta un nivel de 17.5%, pero pasado unos 5 minutos decae espectacularmente hasta perder un 2.5% de rating, aunque el cambio en términos numéricos no es alta, y se mantiene con cambios muy leves hasta 21:21 horas y luego la tendencia experimenta un repunte considerable aproximadamente a las 21:30, aunque sigue sin alcanzar los niveles del inicio del programa, luego nuevamente hay un decaimiento lento hasta un nivel casi como al principio del programa y finalmente termina en un nivel en el orden de 16.6%. En realidad, éste es el programa más regular porque la variación durante el desarrollo de todo el programa es apenas de un 2% en el rating.

El noticiero MEGANOTICIAS, a diferencia de los anteriores presenta en general una tendencia a subir, aunque se mantiene siempre por debajo de ellas, su rating varía desde 9 a 12 por ciento, en la primera media hora del programa, el comportamiento de la tendencia es de subir por unos momentos, aunque la pérdida en las bajas es menos que cuando sube el rating, finalmente la baja más fuerte se produce mas o menos a las 21:31 horas, para luego repuntar y alcanzar un nivel de hasta 11.73% terminando exitosamente el programa, pero siempre manteniendose por debajo de de los otros dos noticieros competidores.

Los resultados en cuanto a los residuos de predicción en los tres casos muestran que el modelo

propuesto se ajusta razonablemente bien a los datos y las predicciones en algunos puntos donde no hay cambios bruscos resulta acorde con lo esperado, aunque estas predicciones no deberian hacerse con horizontes grandes. La metodología propuesta como en el modelo Poisson, abre la posibilidad de modelar series de tiempo para observaciones provenientes de la distribución Binomial o con algunas extensiones menores de observaciones de otras distribuciones pertenecientes a la familia exponencial.

Apéndice A

CALCULOS AUXILIARES

A.1 Momentos en el Modelo Poisson–Gama

Aquí hacemos la deducción de los momentos para en parámetro natural en el modelo Poisson–Gama, donde la función de enlace es el log, que relaciona la media de la variable Poisson con el parámetro citado anteriormente. En efecto

$$Y|\mu \sim Poisson(\mu) \qquad y \qquad \mu \sim Gama(r,s)$$

donde el parámetro natural está dado por

$$\eta = \log \mu$$

en estas condiciones la distribución de probabilidades para μ tiene densidad

$$p(\mu) = \frac{s^r}{\Gamma(r)}\mu^{r-1}e^{-s\mu}, \mu > 0 \tag{A.1}$$

con $r > 0$, $s > 0$ y $\mu > 0$; luego

$$\Gamma(r) = \int_0^\infty s^r \mu^{r-1} e^{-s\mu} d\mu.$$

Derivando esta identidad con respecto a r, tenemos

$$\begin{aligned} \dot{\Gamma}(r) &= \int_0^\infty s\log(s\mu)(s\mu)^{r-1}e^{-s\mu}d\mu \\ &= \int_0^\infty (\log s + \log \mu)s^r\mu^{r-1}e^{-s\mu}d\mu \end{aligned}$$

luego

$$\frac{\dot{\Gamma}(r)}{\Gamma(r)} = \int_0^\infty \log\mu \frac{s^r\mu^{r-1}}{\Gamma(r)}e^{-s\mu}d\mu + \log s \int_0^\infty \frac{s^r\mu^{r-1}}{\Gamma(r)}e^{-s\mu}d\mu$$

Así

$$\gamma(r) = \mathbb{E}(\log \mu) + \log s$$

puesto que la última integral es **1**. Entoces

$$\mathbb{E}(\eta) = \gamma(r) - \log s. \tag{A.2}$$

Para deducir la varianza, derivamos nuevamente la Ecuación (A.2) respecto a r

$$\begin{aligned}
\dot{\gamma}(r) &= \int_0^\infty (\log \mu) s \, \frac{\log(s\mu)(s\mu)^{r-1}\Gamma(r) - \dot{\Gamma}(r)(s\mu)^{r-1}}{\Gamma(r)^2} \, e^{-s\mu} d\mu \\
&= \int_0^\infty \log \mu \, \frac{\log(s\mu)\Gamma(r) - \dot{\Gamma}(r)}{\Gamma(r)^2} \, s^r \mu^{r-1} e^{-s\mu} d\mu \\
&= \int_0^\infty (\log \mu)^2 \frac{s^r \mu^{r-1}}{\Gamma(r)} e^{-s\mu} d\mu \; + \; \log s \int_0^\infty \frac{\log s}{\Gamma(r)} s^r \mu^{r-1} e^{-s\mu} d\mu \\
&\quad -\gamma(r) \int_0^\infty \log \mu \frac{s^r \mu^{r-1}}{\Gamma(r)} e^{-s\mu} d\mu \\
&= \mathbb{E}(\log^2 \mu) \; + \; [\log s - \gamma(r)]\mathbb{E}(\log \mu) \\
&= \mathbb{E}(\log^2 \mu) - \mathbb{E}^2(\log \mu) \\
&= \mathbb{V}(\log \mu) \; = \; \mathbb{V}(\eta).
\end{aligned}$$

En el contexto dinámico, los resultados anteriores para los momentos de la distribución a priori de η_t se pueden escribir como

$$\begin{aligned}
\mathbb{E}(\eta_t | D_{t-1}) &= \gamma(r_t) - \log s_t \\
\mathbb{V}(\eta_t | D_{t-1}) &= \dot{\gamma}(r_t)
\end{aligned}$$

donde $\gamma(\cdot)$ es la función *digamma*, definida para $x > 0$ por

$$\gamma(x) = \frac{d \log \Gamma(x)}{dx}.$$

Generalmente para hacer evaluaciones de esta función, se recurre a su expansión en serie de potencias, los cuales hoy en día estan implementadas en las rutinas computacionales, los cuales estan disponibles en la red *Internet* para cálculos de alta precisión.

A.2 Momentos en el Modelo Beta–Binomial

En esta sección, presentamos las deducciones de los momentos para el parámetro natural, ahora relacionados con el modelo Beta–Binomial, donde la función de enlace es el *logit*, que relaciona la probabilidad de éxito de la binomial con el parámetro citado anteriormente. En efecto, para n conocido

$$Y|\theta \sim Bin(n,\theta), \qquad \theta \sim Beta(r,s).$$

En este caso el parámetro natural, $\eta = logit(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$

En estas condiciones, la densidad de θ es

$$p(\theta) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)}\theta^{r-1}(1-\theta)^{s-1}, \qquad 0<\theta<1 \tag{A.3}$$

dado que $\int_0^1 p(\theta)d\theta = 1$, se tiene que

$$\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 \theta^{r-1}(1-\theta)^{s-1}d\theta. \tag{A.4}$$

Ahora derivando la identidad anterior, respecto a r y luego respecto a s, posteriormente haciendo la diferencia entre ellos, tenemos

$$\begin{aligned}
\frac{\partial}{\partial r}: \quad & \frac{\dot\Gamma(r)\Gamma(r+s)\Gamma(s) - \dot\Gamma(r+s)\Gamma(r)\Gamma(s)}{\Gamma^2(r+s)} &= \int_0^1 \log\theta\ \theta^{r-1}(1-\theta)^{s-1}d\theta \\
\frac{\partial}{\partial s}: \quad & \frac{\dot\Gamma(s)\Gamma(r+s)\Gamma(r) - \dot\Gamma(r+s)\Gamma(r)\Gamma(s)}{\Gamma^2(r+s)} &= \int_0^1 \log(1-\theta)\theta^{r-1}(1-\theta)^{s-1}d\theta.
\end{aligned}$$

Ahora hallamos la diferencia del primero con el segundo

$$\begin{aligned}
\frac{\dot\Gamma(r)\Gamma(s)}{\Gamma(r+s)} - \frac{\dot\Gamma(s)\Gamma(r)}{\Gamma(r+s)} &= \frac{\dot\Gamma(r)}{\Gamma(r)}B(r,s) - \frac{\dot\Gamma(s)}{\Gamma(s)}B(r,s) \\
&= \int_0^1 \log\left(\frac{\theta}{1-\theta}\right)\theta^{r-1}(1-\theta)^{s-1}d\theta
\end{aligned}$$

$$\gamma(r) - \gamma(s) = \int_0^1 logit(\theta)\frac{\theta^{r-1}}{B(r,s)}(1-\theta)^{s-1}d\theta = \mathbb{E}(\eta)$$

donde

$$B(r,s) = \frac{\dot\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

De manera análoga como en el caso anterior, obteniendo las segundas derivadas y después de algunas simplificaciones resulta

$$\mathbb{V}(\eta) = \dot\gamma(r) + \dot\gamma(s).$$

Por lo tanto, en el contexto dinámico, los momentos de la distribución a priori para η_t se escriben

$$\begin{aligned}
\mathbb{E}(\eta_t|D_{t-1}) &= \gamma(r_t) - \gamma(s_t) \\
\mathbb{V}(\eta_t|D_{t-1}) &= \dot\gamma(r_t) + \dot\gamma(s_t).
\end{aligned}$$

A.3 Momentos en el Modelo Binomial Negativa–Beta

Finalmente, deducimos los momentos para el parámetro natural en el modelo el modelo Binomial Negativa–Beta. Sea

$$Y|\mu \sim BinNeg(p,\mu), \qquad \mu \sim Beta(r,s).$$

En este caso, el parámetro natural está dado por

$$\eta = \log(1-\mu).$$

Por lo tanto siguiendo la misma idea como en el caso Binomial, pero en este caso solamente derivando la identidad (A.4) respecto a s, tenemos

$$\frac{\partial}{\partial s}: \qquad \frac{\dot{\Gamma}(s)\Gamma(r+s)\Gamma(r) - \dot{\Gamma}(r+s)\Gamma(r)\Gamma(s)}{\Gamma^2(r+s)} = \int_0^1 \log(1-\mu)\mu^{r-1}(1-\mu)^{s-1}d\mu$$

luego con $B(r,s)$ definida como antes, queda

$$\begin{aligned}
\gamma(s)B(r,s) - \gamma(r+s)B(r,s) &= \int_0^1 \log(1-\mu)\mu^{r-1}(1-\mu)^{s-1}d\mu \\
\gamma(s) - \gamma(r+s) &= \int_0^1 \log(1-\mu)\frac{\mu^{r-1}(1-\mu)^{s-1}}{B(r,s)}d\mu.
\end{aligned}$$

Por lo tanto

$$\mathbb{E}(\eta = \gamma(s) - \gamma(r+s).$$

Derivando nuevamente la última ecuación anterior respecto a s, tenemos

$$\begin{aligned}
\dot{\gamma}(s) - \dot{\gamma}(r+s) &= \int_0^1 \log(1-\mu)\mu^{r-1}\left(\frac{\log(1-\mu) - \gamma(s) + \gamma(r+s)}{B(r,s)}\right)(1-\mu)^{s-1}d\mu \\
&= \int_0^1 \log^2(1-\mu)\frac{\mu^{r-1}(1-\mu)^{s-1}}{B(r,s)}d\mu \\
&\quad -[\gamma(s) - \gamma(r+s)]\int_0^1 \log(1-\mu)\frac{\mu^{r-1}(1-\mu)^{s-1}}{B(r,s)}d\mu \\
&= \mathbb{E}(\eta^2) - \mathbb{E}^2(\eta) = \mathbb{V}(\eta).
\end{aligned}$$

Luego, en términos dinámicos los momentos de la distribución a priori para η_t son

$$\begin{aligned}
\mathbb{E}(\eta_t|D_{t-1}) &= \gamma(s_t) - \gamma(r+s) \\
\mathbb{V}(\eta_t|D_{t-1}) &= \dot{\gamma}(s_t) - \dot{\gamma}(r+s)
\end{aligned}$$

Apéndice B

SUBRUTINAS COMPUTACIONALES EN S–PLUS

La mayoría de los cálculos y gráficas que se presentan en este trabajo fueron desarrollados en el lenguaje–software estadístico S–plus, Por su complejidad y rapidez en la evaluación se rescató un programa fortran llamada `psifn.f` que está disponible en la red *Internet* y fue implementado con un pequeño programa en S–plus para hacer la evaluación de la función digamma definida en la Sección 2.5. Para hacer comparaciones de la bondad de nuestros modelos también se ajustaron los datos con otro software computacional que está diseñado para ajustar modelos estructurales bajo el supuesto de normalidad, llamado STAMP. En particular con los datos disponibles, el ajuste como una aproximación Normal para las observaciones no fueron muy satisfactorios en comparación a lo hecho en este trabajo.

Por otra parte, todas las rutinas estan diseñadas para ajustar modelos dinámicos de tipo Poisson y Binomial para las observaciones, sin embargo, de manera inmediata se puede implementar para otro tipo de observaciones en la familia exponencial o incluso para modelos no lineales, en ese caso simplemente se tendrá que aumentar más opciones en todas las rutinas donde sea necesario.

B.1 Rutinas Generales

El comando que hace posible el uso de la rutina ejecutable `psifn.o` cuyo programa principal está escrita en fortran llamada `psifn.f` el cual puede ser rescatada de la red *internet* o escribir vía *e–mail* a `psunagua@poincare.mat.puc.cl` o `porfirio@cmat.bo`

```
dyn.load2("psifn.o")
```

La matriz G_t del modelo en la Definición 2.2.1

```
G <- function(t) matrix(c(1,0,1,1),2,2)
```

La matriz de Coeficientes del modelo en la Definición 2.2.1

```
Ft <- function(tp) as.matrix(c(1,0))
```

Una variable *dummy*, se puede incorporar al modelo, en caso de observar cambios repentinos que no pueda capturar la tendencia usual del modelo, sin embargo para los modelos ajustados en las aplicaciones se supone que no hay esos comportamientos, por eso definimos solamente como **0**.

```
dummy <- function(tp) 0
```

La función rutina.rs entrega las soluciones de la Ecuación (2.13) dando las estimaciones de los parámetros r_t y s_t de la distribución a priori conjugada definida en la Sección 2.1 para cada familia y función de enlace espesificada, con el cual su distribución de probabilidades queda completamente determinada.

```
rutina.rs <- function(f,q,fam,link,tp)
{
 r <- switch(fam,
        Poisson=switch(link,
                Id =f^2/q,
                log=trigamma.inv(q) ),
        Binomial=switch(link,
                Id =f*(f*(100-f)/q - 1)/100,
                logit=(1+exp(f))/q ) )
 if(r<=0) stop("Parametro r negativo")
 s <- switch(fam,
        Poisson=switch(link,
                Id=f/(q*nt(tp)),
                log=exp(digamma(r)-f)/nt(tp) ),
        Binomial=switch(link,
                Id=(100-f)*(f*(100-f)/q - 1)/100,
                logit=(1+exp(-f))/q ) )
 if(s<=0) stop("Parametro r o s negativo")
 list(r=r,s=s)
}
```

La función UPDATE permite encontrar la media y la varianza a posteriori en el proceso del análisis conjugado descritas en las Secciónes 1.2.2 y 2.3 en el proceso de filtro de Kalman, siempre para cada especificación de la familia y función de enlace especificada.

```
UPDATE <- function(r,s,y,fam,link,tp)
{
 n <-  nt(tp)
 f1 <- switch(fam,
        Poisson=switch(link,
                Id=(r+y)/(n*(s+1)),
                log=digamma(r+y)-log(n*(s+1)) ),
```

```
        Binomial=switch(link,
                Id=100*(r+y)/(r+s+n),
                logit=digamma(r+y) - digamma(s + n - y) )
      )
 q1 <- switch(fam,
        Poisson=switch(link,
                Id=f1/(n*(s+1)),
                log=trigamma(r+y) ),
        Binomial=switch(link,
                Id=f1*(100-f1)/(r+s+n+1),
                logit=trigamma(r+y) + trigamma(s + n - y) )
        )
list(f1=f1,q1=q1)
}
```

El cálculo de la probabilidad para la predicción a 1-paso es entregado por la función prob.p, el mismo que sirve para evaluar la función de verosimilitud, Ecuación (2.47); o para evaluar las probabilidades de predicción en el caso del número de goles del modelo Poisson, Ecuación (3.18).

```
prob.p <- function(y,r,s,fam,tp)
switch(fam,
   Poisson={
       p <- s/(s+1)
       gamma(r+y)/(gamma(r)*gamma(y+1)) * p^r * (1-p)^y
   },
   Binomial={
       n  <- nt(tp)
       exp(lgamma(r+s)+lgamma(n+1)+lgamma(r+y)+lgamma(s+n-y)-
       (lgamma(r)+lgamma(s)+lgamma(r+s+n)+lgamma(n-y+1)+lgamma(y+1)))
   } )
```

La estimación de los residuos de Pearson definida en la Sección 2.8, es proporcionada por la siguiente función r.pearson, que además entrega una posible predicción para el caso de datos faltantes en el cálculo de probabilidades de predicción necesaria en el cálculo de la "devianza".

```
r.pearson <- function(r,s,fam,yt,tp)

 yp <- switch(fam,
         Poisson=r/s,
         Binomial=nt(tp)*r/(r+s) )
 Vp <- switch(fam,
         Poisson=(s+1)*r/s^2,
         Binomial=nt(tp)*r*s*(r+s+nt(tp))/((r+s)^2*(r+s+1)) )
 if(!missing(yt)) rp <- (yt - yp)/sqrt(Vp) else rp <- 0
 list(rp=rp,yp=yp)
```

La siguiente función `s.trend` una vez calculadas las componentes suavizadas $a_t(k)$ y $R_t(k)$ de la Sección 2.4 entrega especificamente el nivel de la tendencia suavizada más su error estándar.

```
s.trend <- function(a,R,tp,d,fam,link)
{
      a <- as.matrix(a);  R <- as.matrix(R)
      E <- Ft(tp)
      f <- c(t(E) %*% a) + d[length(d)]*dummy(tp)
      q <- c(t(E) %*% R %*% E)
      np <- nt(tp)
      if(link!="Id"){
            rs <- rutina.rs(f,q,fam,link,tp)
            r  <- rs$r
            s  <- rs$s
      }
      sn <- switch(fam,
            Poisson = switch(link,Id=a[1,],log=r/(s*np) ),
            Binomial= switch(link,Id=a[1,],logit=100*r/(r+s) )
            )
      er <- switch(fam,
            Poisson=switch(link,Id=sqrt(R[1,1]),log=sqrt(r)/(s*np) ),
            Binomial=switch(link,
                    Id = sqrt(R[1,1]),
                    logit=sqrt(sn*(100-sn)/(r+s+1)) )
            )
      list(snivel=sn,std.e=er)
}
```

La estimación de la tendencia para los datos originales tales como promedio de goles o proporción de rating de la televisión, que no es exactamente la tendencia del modelo estructural que nos proporciona el filtro de Kalman, entonces esta función `Trend` una vez estimada las componentes m_t y C_t para $t = 1, 2, 3, \ldots, T$ de la Sección 2.3, permite calcular la tendencia en línea objetivamente requerida.

```
Trend <- function(r,s,y,C,m,fam,link)
{
    m <- as.matrix(m);  C <- as.matrix(C)
    dy <- length(y)
    z <- y
    np <- nt(1:dy)
    z[is.na(z)] <- switch(fam,
          Poisson=r[is.na(z)]/s[is.na(z)],
          Binomial=np[is.na(z)]*r[is.na(z)]/(r[is.na(z)]+s[is.na(z)])
          )
```

```
    nivel <- switch(fam,
          Poisson =switch(link,
             Id = m[1,],
             log= (r+z)/((s+1)*np) ),
          Binomial= 100*(r+z)/(r+s+np)
          )
    esniv <- switch(fam,
          Poisson =switch(link,
             Id = sqrt(C[1,nrow(C)*(0:(dy-1))+1]),
             log= sqrt(r+z)/((s+1)*np) ),
          Binomial=sqrt(nivel*(100-nivel)/(r+s+np+1))
          )
    list(nivel=nivel,std.e=esniv)
}
```

La función ART es una función auxiliar de la función FCAST, lo que hace es simplemente evaluar $a_t(1)$ y $R_t(1)$ de la Ecuación (2.23).

```
ART <- function(tp,mt,Ct,Wt)
{
        mt <- as.matrix(mt)
        Ct <- as.matrix(Ct)
        Wt <- as.matrix(Wt)
        G1 <- G(tp+1)
        a <- G1 %*% mt
        R <- G1 %*% Ct %*% t(G1) + Wt
        list(a=a,R=R)
}
```

La función FCAST está diseñada especialmente para la predicción a k–pasos para las componentes $m_t(k) = a_t(k)$ y $C_t(k) = R_t(k)$ en el caso de datos faltantes, de este modo se puede continuar con las estimaciones en el fitro de Kalman, ya que el filtro no admite observaciones faltantes.

```
FCAST <- function(tp,K=1,mt,Ct,Wk,d,fam,link)
{
     d  <- as.double(d);  ak <- as.matrix(mt);   Rk <- as.matrix(Ct)
     rk <- sk <- NULL
     for(k in 1:K){
          ar <- ART(tp,ak,Rk,Wk/k)
          ak <- ar$a
          Rk <- ar$R
          E  <- Ft(tp+k)
          fk <- c( t(E) %*% ak) + d[length(d)]*dummy(tp+k)
          qk <- c( t(E) %*% Rk %*% E )
          sr <- rutina.rs(fk,qk,fam,link,tp+k)
```

```
            rk <- append(rk,sr$r)
            sk <- append(sk,sr$s)
        }
        ryp <- r.pearson(rk,sk,fam,,tp+k)
        prk <- prob.p(ryp$yp,rk,sk,fam,tp+k)
        uk <- switch(fam,Poisson=rk/sk,Binomial=100*rk/(rk+sk))
        list(m=ak,C=Rk,a=ak,R=Rk,r=rk,s=sk,prob=prk,rp=0,uk=uk)
}
```

La siguiente función FKALMAN, es una de la rutinas más importantes, que permite hacer estimaciones en línea aplicando el algorítmo de filtro de Kalman. La rutina recibe los valores iniciales m_{t-1} y C_{t-1} y entrega entre otros m_t y C_t para el cuál debe existir la observación y_t.

```
FKALMAN <- function(y,m,C,d,F1,G1,fam,link,tp)
{
        d <- as.double(d);   m <- as.matrix(m);   C <- as.matrix(C)

        a <- G1 %*% m
        if(nrow(C)>1)
            W <- switch(fam,
                      Poisson =diag((1/d[1:nrow(C)]-1)*diag(C)),
                      Binomial=diag(d[1:nrow(C)]) )
        else W <- (1/d-1)*C
        R <- G1 %*% C %*% t(G1) + W

        f <- c( t(F1) %*% a) + d[length(d)]*dummy(tp)
        q <- c( t(F1) %*% R %*% F1)

        rs <- rutina.rs(f,q,fam,link,tp)
        r <- rs$r
        s <- rs$s

        fq <- UPDATE(r,s,y,fam,link,tp)
        f1 <- fq$f1 + d[length(d)]*dummy(tp)
        q1 <- fq$q1

        m <- a + (R %*% F1)*(f1-f)/q
        C <- R - (R %*% F1 %*% t(F1) %*% R)*(1 - q1/q)/q

        pr <- prob.p(y,r,s,fam,tp)

        rp <- r.pearson(r,s,fam,y,tp)$rp
        list(a=a,R=R,m=m,C=C,W=W,r=r,s=s,prob=pr,rp=rp)
}
```

Esta es otra de las funciones importantes de este conjunto de rutinas. DGLM permite la aplicación de filtro de Kalman para cada tiempo $t = 1, 2, \ldots, T$, en el caso de datos faltantes usa la función FCAST, cuyas estimaciones los guarda en matrices o vectores según sea el tipo del objeto. Después de todas las iteraciones calcula entre otras cosas la devianza como $-2\log(L)$, donde $L(\cdot)$ es la función de verosimilitud para los hiperparámetros. Si argumento svto es verdadero, la rutina ejecuta la función SMOOTH, el cuál hace el suavizamiento para las componentes estimadas en línea.

```
DGLM <- function(y,m0,C0,d,fam,link,svto=F,yp)
{
   m <- as.matrix(m0)
   C <- as.matrix(C0)
   dm <- nrow(m)
   a <- R <- r <- s  <- pr  <- rp <- NULL
   N  <- length(y)
   for(t in 1:N){
      blok <- (t-1)*dm + 1:dm
      if(!is.na(y[t])){
         fk <- FKALMAN(y[t],m[,t],C[,blok],d,Ft(t),G(t),fam,link,t)
         W  <- fk$W
      }
      else fk <- FCAST(t-1,1,m[,t],C[,blok],W,d,fam,link)

        m  <- as.matrix( cbind(m,fk$m) )
        C  <- as.matrix( cbind(C,fk$C) )
        r  <- append(r,fk$r)
        s  <- append(s,fk$s)
        a  <- as.matrix( cbind(a,fk$a) )
        R  <- as.matrix( cbind(R,fk$R) )
        pr <- append(pr,fk$prob)
        rp <- append(rp,fk$rp)
   }
dev <- -2 * sum(log(pr))
 if(svto){
      pro <- switch(fam,
           Poisson=PPRED(N,N,1,m[,N+1],C[,blok+dm],W,d,fam,link,yp))
      m <- m[,-1]
      C <- C[,-(1:dm)]
      ypm <- switch(fam,Poisson=1,Binomial=100)*y/nt(1:length(y))
      trends <- Trend(r,s,y,C,m,fam,link)
      nivel <- trends$nivel
      std.niv <- trends$std.e
      if(!is.matrix(m)) m <- t(as.matrix(m))
      if(!is.matrix(C)) C <- t(as.matrix(C))
```

```
      sar <- SMOOTH(m,C,a,R,N,N-1,d,fam,link)
      snv <- sar$snivel
      esn <- sar$std.e
      Sak <- sar$ak
      switch(fam,
        Poisson={
           if(nrow(m)>1) pend  <- m[2,]
              else pend <- rep(0,N)
              if(nrow(Sak)>1) spd <- Sak[2,]
              else spd <- rep(0,N)
              list(m=m,C=C,r=r,s=s,a=a,R=R,prob=pr,rp=rp,nivel=nivel,
              std.niv=std.niv,pend=pend,ypm=ypm,snivel=snv,std.snv=esn,
           spend=spd,pr.pred=pro,dev=dev,m0=m0,C0=C0,f.desc=d,link=link)
           },
        Binomial=list(m=m,C=C,r=r,s=s,a=a,R=R,prob=pr,rp=rp,nivel=nivel,
           std.niv=std.niv,snivel=snv,std.snv=esn,S=Sak,ypm=ypm,dev=dev,
           m0=m0,C0=C0,f.desc=d,link=link)
        )

    }
   else list(dev=dev)
}
```

La siguiente función SMOOTH una vez realizada las estimaciones en línea para las distintas componentes del modelo, la función hace las estimaciones suavizadas para las mismas componentes y entrega la tendencia suavizada, su error estándar y la componente $a_t(-k)$ de la Sección 2.6.

```
SMOOTH <- function(m,C,a,R,tp,K,d,fam,link)
{
     m <- as.matrix(m)
     C <- as.matrix(C)
     a <- as.matrix(a)
     R <- as.matrix(R)
     cm <- ncol(m)
     if(missing(tp)) tp <- cm
     if(missing(K)) K <- tp-1
     if(K>=tp) stop("Numero de rezagos debe ser < long. de la serie")
     dm <- nrow(m)
     blok <- (tp-1)*dm + 1:dm
     Ak <- ak <- as.matrix(m[,tp])
     Rk <- as.matrix(C[,blok])
     trd <- s.trend(ak,Rk,tp,d,fam,link)
     sn <- trd$snivel
     std.e <- trd$std.e
```

```
    for(k in 1:K){
        Gk <- G(tp-k+1)
        DR <- eigen(R[,blok-(k-1)*dm])
        IR <- DR$vectors %*% diag(1/DR$values,dm) %*% t(DR$vectors)
        B <- C[,blok-k*dm] %*% Gk %*% IR
        ak <- m[,tp-k] + B %*% (ak - a[,tp-k+1])
        Rk <- C[,blok-k*dm]-B %*% (R[,blok-(k-1)*dm]-Rk) %*% t(B)
        Ak <- as.matrix(cbind(ak,Ak))
        strend <- s.trend(ak,Rk,tp-k,d,fam,link)
        sn <- append(strend$snivel,sn)
        std.e <- append(strend$std.e,std.e)
    }
    list(ak=Ak,snivel=sn,std.e=std.e)
}
```

El cálculo de la primera y segunda derivada para la función digamma $\gamma(\cdot)$ definida en la Sección 2.5, lo hace la función `trigamma`, utilizando para ello la aproximación de su expanción polinomial. Es decir, hace la evaluación de $\dot{\gamma}(\cdot)$ o de $\ddot{\gamma}(\cdot)$.

```
trigamma <- function(x,n=1)
{
  if(x<=0) stop("Argumento de trigamma negativo o cero")
  else
  switch(n,1/x+1/(2*x^2)+1/(6*x^3)-1/(3*x^5)+1/(3*x^5)+1/(42*x^7)
           -1/(30*x^9) + 5/(66*x^11) ,
  -1/x^2-1/x^3-1/(2*x^4)+1/(6*x^6)-1/(6*x^8)+3/(10*x^10)-5/(6*x^12),
  stop("no esta implementado mas derivadas") )
}
```

La función `digamma.inv` permite calcular la inversa de la función digamma, $\gamma(\cdot)^{-1}$, mediante el algoritmo de Newton-Rapson, en donde se usa la función `trigamma` para obtener la primera derivada $\dot{\gamma}(\cdot)$.

```
digamma.inv <- function(f)
{
        err <- 1e-6
        dif <- err + 1
        r0  <- exp(f)
        while(dif>=err){
                r <- r0 - (digamma(r0)-f)/trigamma(r0)
                dif <- abs(r-r0)
                r0 <- r
        }
        r
}
```

Esta función `trigamma.inv` calcula la inversa de la primera derivada de la función digamma, es decir evalúa $\dot{\gamma}(\cdot)^{-1}$.

```
trigamma.inv <- function(q)
{
        err <- 1e-6
        dif <- err + 1
        r0 <- (1+ sqrt(1+2*q))/(2*q)
        while(dif >= err) {
              r <- r0 - (trigamma(r0) - q)/trigamma(r0,2)
              dif <- abs(r-r0)
              r0 <- r
        }
        r
}
```

Esta función para fines de estabilidad numérica del algoritmo del filtro de Kalman, lo que hace es calcular un valor coherente para los valores iniciales del nivel de la tendecia del modelo estructural en términos de una media de los primeros 5 observaciones afectada luego por la función del enlace.

```
Inicial <- function(y,fam,link)
{
      m0 <- mean(y[1:5]/nt(1:5))
      m0 <- m0*switch(fam,Poisson=1,Binomial=100)
      switch(fam,
             Poisson=switch(link,Id = m0,log = log(m0)),
             Binomial=switch(link,Id = m0,logit= log(m0/(100-m0)))
            )
}
```

La función pause posibilita dar una pausa de tiempo para observar gráficos consecutivos en una ventana de gráficas `X11()` de S-plus.

```
pause <- function()
{
        restart(T)
        cat("Enter expresions, q to quit from pause\n")
        repeat {
                e <- parse(prompt = "<P>")
                if(is.name(e[[1]]) && e[[1]] == "q")
                        return()
                print(eval(e, local = sys.parent(1)))
        }
}
```

La siguiente función bautizada como digamma hace la evaluación de la función digamma $\gamma(\cdot)$, el cual es la implementación de la programa Fortran psifn.f que está disponible en la red *internet* cuyo programa original no lo incluimos aquí por que consta aproximadamente de 40 páginas.

```
digamma <- function(x)
{
      cn <- names(x)
      if(is.matrix(x)) {
         rn <- row.names(x)
         d  <- dim(x)
      }
      y <- c(x)
      w <- NULL
      for(i in 1:length(x)){
         z <- .Fortran("psifn",X=as.double(y[i]),
                        n=as.integer(0),
                        KODE=as.integer(2),
                        M=as.integer(1),
                        ANS=as.vector(1),
                        NZ=as.integer(0),
                        IERR=as.integer(0))
         if(z$IERR==0 & z$NZ==0)
             w <- append(w,z$ANS)
         else {if(z$NZ!=0) stop("Underflow, last NZ ANS set to 0")
                  if(z$IERR==1) stop("Input Error, no computation")
                  if(z$IERR==2) stop("Overflow or N+M-2 too large")
                  if(z$IERR==3) stop("N too large")
               }
         }
      if(is.matrix(x))
              w <- - matrix(w,d[1],d[2],dimnames=list(rn,cn))
      else {
           w <- - w
           names(w) <- cn
      }
      w <- w + log(x)
      w
}
```

B.2 Rutinas Particulares Para el Modelo Poisson–Gama

La función nt que permite obtener en número de partidos jugados en cada año, donde uc es un objeto S–plus el la cual está almacenado los datos de los goles y número de partidos.

```
nt <- function(tp)
{
     n <- uc$n
     ifelse(0<tp&tp<=length(n),n[tp],1)
}
```

La función CDGLM ejecuta la rutina DGLM para diversas combinaciones de hiperparámetros, que en este caso son los factores de descuento, calculando para cada una de ellas su "devianza" el cuál se muestra en una tabla juntamente con las estimaciones de probabilidades de predicción usando toda la información disponible para el número de goles que puede ser marcado en el próximo partido del mismo torneo del siguiente año y al final entrega el ajuste del modelo con la devianza óptima. La rutina también entrega, si el argumento graf=TRUE, una ilustración gráfica ocupando la rutina GRAF donde se puede observar visualmente distinatas componentes estimadas.

```
CDGLM <- function(y,D,m0,C0,fam="Poisson",link="log",graf=F,titulo="",
                  year=1939:1994,yp=0:5)
{
     if(missing(D)) D <- matrix(c(.97,.9,.99,.9,.97,.95),2,3)
     if(missing(yp)) yp <- 0:5
     if(missing(m0)) m0 <- as.matrix(c(2,0))
     if(missing(C0)) C0 <- diag(2)
     if(!is.matrix(D)) D <- t(as.matrix(D))
     cd <- ncol(as.matrix(D))
     Dev <- pro <- NULL
     sw <- T
     if(cd>1) sw<-!sw
     for(i in 1:cd){
             sal <- DGLM(y,m0,C0,D[,i],fam,link,svto=sw,yp)
             Dev <- append(Dev,sal$dev)
     }
     if(!sw) sal<-DGLM(y,m0,C0,D[,Dev==min(Dev)],fam,link,svto=T,yp)

     prob <- sal$pr.pred
     dif <- append(0,-diff(Dev) )
     Delta <- t(D)
     Prox.Partido <-  c(prob,1-sum(prob) )

     cat(titulo,fill=T)
     print(data.frame(Delta,Dev,dif))
```

```
    cat("Delta Optimo\t=",c(" ",D[,Dev==min(Dev)]),fill=T)
    Goles <- c(0:(length(prob)-1),paste(">=",length(prob),sep=""))
    print(data.frame(Goles,Prox.Partido))
    if(graf) GRAF(sal,year,yp)
    sal
}
```

La rutina GRAF, muestra la gráfica de las distintas componentes más sobresalientes del modelo, tales como: La *Tendencia en Línea* y *Suavizada*, presentando como fondo los promedio de goles por partido, los *Residuos de Pearson* y finalmemente presenta las gráficas de las estimaciones de probabilidades para el número de goles que puede ser marcado en el próximo partido con información hasta el año de su mejor campaña, la más baja temporada, con toda la información disponible y con el promedio histórico. La función entrega como resultado las estimaciones de las probabilidades para cada caso.

```
GRAF <- function(sal,year=1939:1994,yp=0:5)
{
    if(dev.set()==1) X11()
    par(mfrow=c(1,1))
    fam <- "Poisson"
    N  <- prox <- length(sal$ypm)
    ylin <- "Promedio de Goles y Tendencia (en linea)"
    plot(year,sal$ypm,main=ylin,xlab="",ylab="")
    lines(year,sal$nivel)
    abline(h=mean(sal$ypm[!is.na(sal$ypm)]),lty=2)
    pause()

    ysua <- "Promedio de Goles y Tendencia (Suavizada)"
    plot(year,sal$ypm,main=ysua,xlab="",ylab="")
    lines(year,sal$snivel,lty=1)
    abline(h=mean(sal$ypm[!is.na(sal$ypm)]),lty=2)
    pause()

    plot(year,sal$rp,main="Residuos de Pearson",xlab="",ylab="",type="l")
    abline(h=0,lty=2)
    pause()

    par(mfrow=c(2,2))
    alto <- (1:N)[sal$nivel==max(sal$nivel)]
    bajo <- (1:N)[sal$nivel==min(sal$nivel)]
    dm <- nrow(sal$m0)
    d1 <- as.matrix(c(sal$f.desc))

    blk <- (alto-1)*dm + 1:dm
```

```
    W <- diag((1/sal$f.desc-1)*diag(sal$C[,blk]))
    Palto <- PPRED(N,alto,1,sal$m[,alto],sal$C[,blk],W,d1,fam,sal$link,yp)
    Prox.Mejor <- c(Palto,1-sum(Palto))

    blk <- (bajo-1)*dm + 1:dm
    W <- diag((1/sal$f.desc-1)*diag(sal$C[,blk]))
    Pbajo <- PPRED(N,bajo,1,sal$m[,bajo],sal$C[,blk],W,d1,fam,sal$link,yp)
    Prox.Malo  <- c(Pbajo,1-sum(Pbajo))

    pprox <- sal$pr.pred
    Prox.Partido <- c(pprox,1-sum(pprox))
    lambda <- mean(sal$ypm[!is.na(sal$ypm)])
    phist  <- dpois(yp,lambda)
    Prob.Hist <- c(phist,1-sum(phist))

    Years <- c(year[alto],year[bajo],year[prox],lambda)
    names(Years) <- c("Mejor","Malo","Futuro","Prom.hist")
    Goles <- c(yp,paste(">=",length(yp),sep=""))
    mat <- data.frame(Goles,Prox.Mejor,Prox.Malo,Prox.Partido,Prob.Hist)
    cat("Probabilidad del proximo partido",fill=T)

    mx <- max(Palto,Pbajo,pprox,phist)
    Pb <- "Probabilidad"
    pp <- "Probabilidad Proximo Partido"
    plot(yp,Palto,ylim=c(0,mx),main=pp,xlab="Goles",ylab=Pb,type="l")
    text(3.5,4*mx/5,paste("En",year[alto],sep=" "))
    plot(yp,Pbajo,ylim=c(0,mx),main=pp,xlab="Goles",ylab=Pb,type="l")
    text(3.5,4*mx/5,paste("En",year[bajo],sep=" "))
    plot(yp,pprox,ylim=c(0,mx),main=pp,xlab="Goles",ylab=Pb,type="l")
    text(3.5,4*mx/5,paste("En",year[prox],sep=" "))
    plot(yp,phist,ylim=c(0,mx),main=pp,xlab="Goles",ylab=Pb,type="l")
    text(3.5,4*mx/5,"Con Promedio")
    print(Years)
    mat
}
```

DIVISION BIENES E INVENTARIOS
VERIFICADO
Fecha
H. M. R
M. S

Esta función PPRED calcula las probabilidades de predicción del número de goles para el próximo partido con información hasta un año especificado.

```
PPRED <- function(N,tp,k=1,mt,Ct,Wk,d,fam,link,yp)
{
    d  <- as.double(d)
    mt <- as.matrix(mt)
    Ct <- as.matrix(Ct)
```

```
      ar <- ART(tp,mt,Ct,Wk)
      ak <- ar$a
      Rk <- ar$R
      E  <- Ft(tp+k)
      fk <- c( t(E) %*% ak) + d[length(d)]*dummy(tp)
      qk <- c( t(E) %*% Rk %*% E )
      sr <- rutina.rs(fk,qk,fam,link,N+k)
      rk <- sr$r
      sk <- sr$s
      prob.p(yp,rk,sk,fam,tp+k)
}
```

Con datos del número de goles por partido tanto de la Universidad de Chile como de la Universidad Católica, además de los años y partidos que corresponden cada resultado, la función numgol entrega entre otros el promedio de goles anuales (pga) en este caso para la U. Católica, el número de partidos por año (n), más detalles se puede encontrar en el propio programa.

```
numgol <- function(pt,year,A,B){
        a  <- A[pt=="1P"]        # No.de goles de la 1ra. Rueda
        b  <- A[pt=="2P"]        # No.de goles de la 2da. Rueda
        c  <- a + b              # Suma de los dos partidos:1ra+2da
        ge <- rep(0,length(a))   # Pr.de pt. ganados o empatados
        pg <- ge                 # Pr.de partidos ganados
        pa <- ge                 # Promedio de goles por a\no
        ne <- ge                 # Num. pt. ganados o empatados
        ng <- ge                 # Num. de partidos ganados
        np <- ge                 # Num. de pt. Efect. Jugados c/a.
        ga <- ge                 # Num. de goles anuales
        n  <- ge                 # Num pt c/a.
        for(i in 1939:1994){
                ge[i-1938]  <- mean(A[year==i]>=B[year==i])
                ne[i-1938]  <- sum(A[year==i]>=B[year==i])
                pg[i-1938]  <- mean(A[year==i]>B[year==i])
                ng[i-1938]  <- sum(A[year==i]>B[year==i])
                pa[i-1938]  <- mean(A[year==i])
                ga[i-1938]  <- sum(A[year==i])
                np[i-1938]  <- sum(!is.na(A[year==i]))
                 n[i-1938]  <- sum(year==i)
                }
list(a1p=a,a2p=b,a12p=c,nge=ne,pge=ge,npg=ng,pg=pg,nga=ga,np=np,pga=pa,n=n)
}
```

B.3 Rutinas Particulares Para el Modelo Binomial–Beta

La función nt provee el valor de n_t para el modelo binomial, que es el tamaño de muestra de hogares televidentes, el cuál en este caso resulta una constante, pero en otros casos podría variar a través del tiempo.

```
nt <- function(tp) 324
```

La función CDGLM cuyo procedimiento es escencialmente lo mismo que para el caso del modelo Poisson, de hecho, se puede usar la misma función evaluando convenientemente en parámetros adecuados, pero por simplicidad en su aplicación presentamos nuevamente la función adaptada para el modelo Binomial, que en escencia hace la evaluación de la devianza para cada combinación dada para los hiperparámetros, entregando al final los resultados de las estimaciones de las componentes con el hiperparámetro óptimo y que alternativamente se puede pedir a la función para que muestre graficamente las estimaciones de la tendencia en línea y suavizada y los residuos de Pearson.

```
CDGLM <- function(y,D,m0,C0,fam="Binomial",link="Id",graf=F,titulo="")
{
      if(missing(D))
          D <- matrix(switch(link,Id=c(.1,1),logit=c(1e-10,.005)),2,1)
      if(missing(m0)) m0 <- as.matrix(c(Inicial(y,fam,link),0))
      if(missing(C0)) C0 <- diag(switch(link,Id=c(.1,.1),logit=c(.01,.01)))
      if(!is.matrix(D)) D <- t(as.matrix(D))
      cd <- ncol(as.matrix(D))
      Dev <- pro <- NULL
      sw <- T
      if(cd>1) sw<-!sw
      for(i in 1:cd){
              sal <- DGLM(y,m0,C0,D[,i],fam,link,svto=sw)
              Dev <- append(Dev,sal$dev)
      }
      if(!sw) sal<-DGLM(y,m0,C0,D[,Dev==min(Dev)],fam,link,svto=T)

      dif <- append(0,-diff(Dev) )
      Delta <- t(D)

      cat(titulo,fill=T)
      print(data.frame(Delta,Dev,dif))
      cat("Delta Optimo\t=",c(" ",D[,Dev==min(Dev)]),fill=T)
      if(graf) GRAF(sal)
      sal
}
```

Entregando la *salida* de la función CDGLM la rutina GRAF para una apresiación visual presenta la gráfica de las estimaciones en línea y suavizada para la tendencia más los datos de los rating y los residuos de Pearson del modelo ajustado.

```
GRAF <- function(sal=stop("Sin datos?"),sin.t=F)
{
        if(dev.set()==1) X11()
        par(mfrow=c(1,1))
        Times <- time(sal$ypm)[1:length(sal$ypm)]
        horas <- as.integer(Times)
        horas <- horas + (Times-horas)*60/100

        rab <- range(c(sal$ypm,sal$nivel))
        tsplot(sal$ypm,type="p",pch="*",xaxt="n",ylim=rab)
        tslines(sal$nivel)
        axis(1,at=time(sal$ypm),lab=horas)
        title(ifelse(sin.t,"","TVR Hogar y Tendencia (En Linea)"))
        pause()

        param <- tsp(sal$nivel)
        tsplot(sal$ypm,type="p",pch="*",xaxt="n")
        axis(1,at=time(sal$ypm),lab=horas)
        tslines(ts(sal$snivel,start=param[1],frequency=param[3]))
        title(ifelse(sin.t,"","TVR Hogar y Tendencia (Suavizada)"))
        pause()

        tsplot(ts(sal$rp,start=param[1],frequency=param[3]),xaxt="n")
        axis(1,at=time(sal$ypm),lab=horas)
        title(ifelse(sin.t,"","Residuos de Pearson"))
        abline(h=0,lty=2)
}
```

La función texto facilita colocar textos en los gráficos en áreas especificadas por el mouse del computador con un tamaño de letra especificado.

```
texto <- function(tam=1)
{
        cat("<texto> ==>")
        txt <- scan(what="",n=1,sep="*")
        cat("Ahora hacer un <click>",fill=T)
        text(locator(1),txt,cex=tam)
}
```

La hora de inicio y finalización de los programas de televisión es proporcionada por la función ini.ter, para el cual la función lee el archivo prog.tv, especificando por el canal de la televisión, día y el nombre del programa.

```
ini.ter <- function(cnl=13,dia="lun",prog="TL13")
{
        hp <- read.table(file=
            "/nfs/poincare/users/mat/psunagua/tesis/rating/prog.tv",T)
        cls <- c("H.INI","M.INI","H.TER","M.TER")
        hp <- hp[hp[,"CANAL"]==cnl&hp[,"DIA"]==dia&hp[,"PROG"]==prog,cls]
        hp <- c(hp[,1]+hp[,2]/100,hp[,3]+hp[,4]/100)
        hp
}
```

La función tv.com permite obtener la predicciones para el rating en los momentos de los comerciales que ocurrieron durante un programa, para ello requiere de la salida del ajuste del modelo con la función CDGLM y del archivo de datos horas.com.

```
tv.com <- function(out,cnl=13,dia="lun",prog="")
{
  rtp <- range(time(out$ypm))
  dm <- nrow(out$m); d  <- out$f.desc
  hc <- read.table(file=
        "/nfs/poincare/users/mat/psunagua/tesis/rating/horas.com",T)
  hc <- hc[hc[,"CANAL"]==cnl&hc[,"DIA"]==dia,-(1:2)]
  hctot <- hc[,"HORA"]+hc[,"MIN"]/60 + hc[,"SEG"]/3600
  hc <- hc[hctot>=rtp[1] & hctot<=rtp[2],]
  W <- diag(out$f.desc[1:nrow(out$C)])
  if(dev.set()==1) X11(); par(mfrow=c(2,2)); mto <- 4
  M <- tpo <- Inicom <- Durac <- NULL
  for(i in 1:nrow(hc)){
        hrc <- hc[i,"HORA"]+hc[i,"MIN"]/60 + hc[i,"SEG"]/3600
        ini <- as.integer((hrc-rtp[1])*60)+1
        fst <- max(1,ini-mto)
        hri <-  hc[i,"HORA"]+(hc[i,"MIN"]-(ini-fst))/60
        dur <- max(2,round(hc[i,"DURACION"]/60))
        blk <- (ini-1)*dm+1:dm
        fc <- FCAST(ini,dur,out$m[,ini],out$C[,blk],W,d,"Binomial",out$link)
        yt <- out$ypm[fst:min((ini+dur),length(out$ypm))]
        uk <- c(out$nivel[ini],fc$uk)
        rgo <- range(c(yt,out$nivel[fst:(ini)],uk))+c(-1,1)/2
        plot(yt,xaxt="n",xlab="",ylab="% TVR",ylim=rgo)
        title(paste(prog," (",dia,")",sep=""),cex=1)
        lines(out$nivel[fst:(ini)])
        lines(ini-fst+1:length(uk),uk,lty=3)
        abline(v=ini-fst+1+hc[i,"SEG"]/60,lty=2)
        hri <- seq(hri,by=1/60,length=dur+mto+1)
        hri <- as.integer(hri)+(hri-as.integer(hri))*60/100
```

```
        fin <- length(hri)
        text((ini-fst+1+hc[i,"SEG"]/60+fin)/2,rgo[2],"Comerciales",cex=.5)
        inicom <- paste(hc[i,"HORA"],":",hc[i,"MIN"],":",hc[i,"SEG"],sep="")
        text(ini-fst+1+hc[i,"SEG"]/60,rgo[1],inicom,cex=.5)
        Inicom <- paste(Inicom,inicom)
        durac <- paste(hc[i,"DURACION"],"seg.")
        text((ini-fst+1+hc[i,"SEG"]/60+fin+.5)/2,rgo[1],durac,cex=.5)
        Durac <- paste(Durac,durac)
        axis(1,at=1:(dur+mto+1),lab=hri)
        ley <- c("Actual","En Linea","Prediccion")
        legend(1,rgo[2],ley,lty=c(0,1,3),pch="*  ",cex=.5)
        per <- pbeta(yt[ini-fst+2:length(uk)]/100,fc$r,fc$s)*100
        if(as.integer(i/4)==i/4 & nrow(hc)>4) pause()
        tpo <- append(tpo,0:(length(uk)-1))
        hrs <- hri[ini-fst+1:length(uk)]
        y <- yt[ini-fst+1:length(uk)]
        M <- rbind(M,cbind(i,hrs,uk,y,c(0,per),y-uk))
}
dimnames(M) <- list(tpo,c("COM","HORA","PRED","OBS","PCTL","ERP"))
cat("Canal:",cnl," Dia:",dia," Programa:",prog,"\nComercial:",Inicom,
    "\nDuracion: ",Durac,fill=T)
M
}
```

La siguiente función read lo único que hace es leer del archivo de datos una columna especificada para los rating y entrega ésta convertida en una serie de tiempo más una gráfica de la serie, el cual es usado para el ajuste del modelo con la función CDGLM.

```
read <- function(archdia,ini=19.30,ter=23,serie="HOG.13",subdir)
{
   if(missing(subdir))
         subdir<-"/nfs/poincare/users/mat/psunagua/tesis/rating/"
   hora <- as.integer(ini)
   minuto <- (ini-hora)*100
   archivo <- paste(subdir,archdia,sep="")
   tserie<-read.table(archivo,T)
   row.names(tserie) <- tserie[,"HORA"]
   yt<-tserie[ini<=tserie[,"HORA"]&tserie[,"HORA"]<=ter,c(serie,"HORA")]
   tsyt <- ts(yt[,serie],start=hora+minuto/60,frequency=60)
   if(dev.set()==1) X11();  par(mfrow=c(1,1))
   tsplot(tsyt,xaxt="n")
   axis(1,at=time(tsyt),lab=yt[,"HORA"])
   324*tsyt/100
```

Finalmente la función pred permite hacer las predicciones con un horizonte especificada en cualquier instante del tiempo monitoriada por la gráfica de la estimación en línea para la tendencia, claro que para esto se requiere de la salida del ajuste del modelo, los puntos de predicción son seleccionadas visualmente con el mouse y al final la rutina entrega entre otros las predicciones, errores de predicción, observaciones actuales y horas.

```
pred <- function(out,dur=4)
{
    if(dev.set()==1) X11();  par(mfrow=c(1,1))
    Times <- time(out$ypm)[1:length(out$ypm)]
    horas <- as.integer(Times)
    horas <- horas + (Times-horas)*60/100
    rab <- range(c(out$ypm,out$nivel))+c(-.5,1)
    tsplot(out$ypm,type="p",pch="*",xaxt="n",ylim=rab)
    tslines(out$nivel)
    axis(1,at=time(out$ypm),lab=horas)
    rtp <- range(time(out$ypm))
    dm <- nrow(out$m); d  <- out$f.desc
    W <- diag(out$f.desc[1:nrow(out$C)])
    M <- tpo <- NULL
    cat("<click> Punto de Prediccion",fill=T)
    hrini <- locator(1)
    if(rtp[1]>hrini$x&hrini$x>rtp[2]) stop("Fuera de rango")
    ley <- c("Observado","En Linea","Prediccion")
    legend(Times[1],rab[2],ley,lty=c(0,1,3),pch="*  ",cex=1)
    while(rtp[1]<=hrini$x&hrini$x<=rtp[2]){
        ini <- as.integer((hrini$x-rtp[1])*60)+1
        blk <- (ini-1)*dm+1:dm
        fc  <- FCAST(ini,dur,out$m[,ini],out$C[,blk],W,d,"Binomial",out$link)
        uk  <- c(out$nivel[ini],fc$uk)
        Time <- Times[ini:(ini+dur)]
        lines(Time,uk,lty=3)
        per <- pbeta(out$ypm[(ini+1):(ini+dur)]/100,fc$r,fc$s)*100
        tpo <- append(tpo,0:dur)
        hrs <- as.integer(Time)+(Time-as.integer(Time))*60/100
        yt  <- out$ypm[ini:(ini+dur)]
        M   <- rbind(M,cbind(hrs,uk,yt,c(0,per),yt-uk))
        cat("<click> Punto de Prediccion(fuera de rango=Terminar)",fill=T)
        hrini <- locator(1)
    }
    dimnames(M) <- list(tpo,c("HORA","PRED","OBS","PCTL","ERP"))
    M
}
```

Bibliografía

[1] Agresti, A. (1990), *Categorical Data Analysis*, John Wiley & Sons, New York.

[2] Aitchison, J. y Dunsmore, I.R.(1975), *Statistical Prediction Analysis*, Cambridge University Press, Cambridge.

[3] Baker, R.J. y Nelder, J.A, (1985), *GLIM Relese 3.77* Oxford University, Numerical Algoritms Group.

[4] Bernardo J. M. y Smith A. F.M.(1994), *Bayesian Theory*, John Wiley & Sons, New York.

[5] Box, G.E.P. y Jenkins, G.M, (1976), *Time Series Analysis: Forecasting y Control*, (2nd. Ed.), Holden–Day, San Francisco.

[6] Brockwell, P. J. y Davis, R. A. (1990), *Time Series: Theory y Methods*, Springer–Verlag, 2nd Edition, New York.

[7] Broadbent S., One Way TV Advertisements Work, *Journal of the Market Research Society*, Volume 21, Number 3, pp. 139–166.

[8] Colman S. y Brown G., Advertising Traking Studies and Sales Effects, *Journal of the Market Research Society*, Volume 25, Number 2, pp. 165–183.

[9] Collet, D. *Modelling Survival Data in Medical Reseach*, Chapman & Hall, London.

[10] De Groot, M.H. (1971), *Optimal Statistical Decisions*, Mc Graw–Hill, New York.

[11] Diggle, P. J.(1990), *Time Series: A Biostatistical Introduction*, Oxford University Press, New York.

[12] Efron, E. y Tibshirani, R.J. (1994), *An Introduction to the Bootstrap*, Chapman & Hall, London.

[13] Fahrmier, L. y Tutz, G.(1994), *Multivariate Statistical Modelling Based on Geleralized Linear Models*, Sringer–Verlag, New York.

[14] Gazmuri, P. (1995) *Modelos Estocásticos para la Gestión de Sistemas*, Ediciones Universidad Católica de Chile, Primera Edición, Santiago de Chile.

[15] Graybill, F. A. *Matrices with Aplications in Statistics*, 2nd. Edition, McGraw Hill, New York.

[16] Harvey, A.C. (1989) *Forecasting, Structural Time Series Models y the Kalman Filter*, Cambridge University Press, Cambridge.

[17] Harvey, A.C. y Fernandes, C. (1989), Time Series Models for Count or Qualitative Observations, *Journal of Business & Economic Statistics*, October Vol. 7 No.4, pp. 407–417.

[18] Marshall, P.(1990), *Modelos Estructurales Para Series Cronológicas*, Cuaderno PUC 92/2, Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile.

[19] Marshall, P.(1993), *Modelos Para Series Cronológicas Aparentemente no Relacionadas*, Cuaderno PUC/FM-93/1, Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile.

[20] Nelder,J. A., y Wedderburn, R. W. M.,(1972), Generalized Linear Models *Journal of the Royal Statistical Society*, Ser. A, 135, pp. 370–384.

[21] McCullagh, P. y Nelder, J.A.,(1989) *Generalized Linear Models*, 2nd. Edition, Chapman & Hall, London.

[22] Pole, A., West, M. y Harrison, J., (1994) *Applied Bayesian Forecasting y Time Series Analysis*, 1st. Edition, Chapman & Hall, London.

[23] West, M. y Harrison, J., (1986), Monotoring y Addaptation in Bayesian Forecasting Models, *Journal of the American Statistical Association*, Vol. 81, No 395, pp. 741–761.

[24] West, M., (1985a), Generalized Linear Models, outlier accomodation , scale parameters y prior distribution. In *Bayesian Statistics 2*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley y A.F.M. Smith (Eds), Valencia University Press.

[25] West, M. y Harrison, J. (1989) *Bayesian Forecasting and Dynamic Models*, Springer–Verlag, New York.

[26] West, M., Harrison, J., y Migon, H.S. (1985), Dynamic Generalized Linear Models y Bayesian Forecasting, *Journal of the American Statistical Association* **80**, pp. 73–97.