

Examen Parcial 2

Elizabeth Solis, Mariana Godina, Andrea Navarrete

Otoño 2017

1. Identifiquen el soporte de las dos sucesiones de variables aleatorias a las que hace referencia el documento, i.e. $\{t_i, m_i\}_{i=1}^N$ y N .

El soporte de la sucesión de la variable aleatoria t_i ocurrencias de sismos es de 0 a ∞ . El soporte de la magnitud del sismo variable aleatoria m_i es $m > 4.7$

2. La distribución de intensidad referida en la ecuación (1.1) es referida a la variable del número de eventos en el intervalo $[0, T]$ (con $T < \infty$), de igual forma la definida por la ecuación (1.2). La función de intensidad o *hazard* se define como la tasa instantánea de cambio de la función de densidad en el tiempo t condicional en que la v.a. es mayor que t . Así, la función de densidad $f(t)$ estaría definida como:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

donde $1 - S(t) = \Pr t \in (0, T]$. Con base en esto, deriva los pasos necesarios para encontrar la expresión (2.1).

De la expresión anterior conocida como *hazard rate* se sigue:

$$\lambda(t) = \frac{f(t)}{S(t)}$$

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

$$\lambda(t) = -\frac{d}{dt} \log[S(t)]$$

$$\implies \frac{d}{dt}[S(t)] = f(t)$$

$$\implies S(t) = e^{-\int_0^T \lambda(t) dt}$$

De esta forma escribimos la similitud de la siguiente forma:

$$L_T(\theta) = \prod_{i=1}^N f(t_i|\theta)$$

$$L_T(\theta) = \prod_{i=1}^N \lambda(t_i) S(t_i)$$

$$\implies L_T(\theta) = \prod_{i=1}^N \lambda(t_i) e^{-\int_0^T \lambda(t) dt}$$

Sacando logaritmo:

$$\implies \log L_T(\theta) = \log \left[\prod_{i=1}^N \lambda(t_i) e^{-\int_0^T \lambda(t) dt} \right]$$

$$\begin{aligned}
\Rightarrow \log L_T(\theta) &= \sum_{i=1}^N \log[\lambda(t_i) e^{-\int_0^T \lambda(t) dt}] \\
\Rightarrow \log L_T(\theta) &= \sum_{i=1}^N \log[\lambda(t_i)] + \log[e^{-\int_0^T \lambda(t) dt}] \\
\Rightarrow \log L_T(\theta) &= \sum_{i=1}^N \log[\lambda(t_i)] - \int_0^T \lambda(t) dt
\end{aligned}$$

NOTA: (Se cambio la función *hazard* por la siguiente función la siguiente, siguiendo la función descrita en el libro: D.J. Daley; D. Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods* (2002), Second Edition. Springer)

3. Con base en el inciso anterior, identifica los parámetros y espacios parametrales correspondientes al modelo dado por (2.1).

El modelo dado por (2.1) es el siguiente:

$$\log L_T(\theta) = \sum_{i=1}^N \log \lambda(t_i; \theta) - \int_0^T \lambda(t; \theta) dt$$

donde $\theta = (\mu, a, \beta, c, p)$ para M1

El espacio parametral es:

- $\mu \in (0, \inf)$
- $p \in [1, \inf)$
- $a > 0$
- $\beta > 0$
- $c > 0$

4. Con base en el inciso anterior, desarrollen interpretaciones para el rol que juegan cada uno de los parámetros del modelo. Para esto, elijan uno entre los modelos M1 y M2 enunciados en el documento.

Con el modelo M1 el rol que juegan cada parámetro es:

$$\lambda(t|\theta) = \mu + \sum_{t_i < t} e^{\beta(m_i - Mr)} \frac{a}{(t - t_i + c)^p}$$

- μ : representa la tasa de actividad o intensidad de fondo
- p : es una tasa de decaimiento de intensidad
- a : es una constante proporcional de la tasa de réplicas
- β : es la tasa de réplicas con magnitudes superiores a un umbral. Si beta es un valor muy chico, quiere decir que se esperan mayores réplicas superiores a un umbral; mientras que si beta es muy grande, habrá pocas réplicas superiores al mismo umbral.
- c : es una constante de decaimiento en el tiempo

5. Con base en la Sección 2.2 del documento, argumenta si los supuestos sobre los cuales los autores proponen la distribución inicial para los parámetros es razonable. Identifica si la distribución inicial es del tipo conjugado para alguno de los parámetros involucrados. Discute la pertinencia de la ecuación (2.4).

Parámetros: $\underline{\theta} = (\mu, a, \beta, c, p)$ μ es independiente de (a, β, c, p) por el hecho de que μ representa los principales shocks y las demás los shocks posteriores. β describe la magnitud del efecto y (c, p) son el tiempo y el efecto de los eventos, por lo tanto se asume independencia entre β y (c, p) . p se distribuye uniforme $(0, 5)$. En cambio, a depende de β, c y p . μ, β y c se distribuyen gamma. a es una distribución gamma con hiperparámetros

que dependen de los valores β , c y p . Estos supuestos son razonables, ya que ocupan información previa para alimntar el conocimiento de la distribución inicial. La distribución a es una distribución conjugada gamma-gamma.

6. Con base en lo visto en clase, y en la Sección 3 del documento, desarrolla un algoritmo para calcular la distribución final de los parámetros. (En este inciso, sólo el diseño del algoritmo y cálculo de la distribuciones condicionales completas es relevante, no la implementación del código).

Dados los datos observados

$$\begin{aligned} \{t_i, m_i\}_{i=1}^N &\in [0, T], \\ m_i &> M_r \forall i = 1, \dots, N \end{aligned}$$

La distribución final o posterior asociada es

$$\pi(\underline{\theta}|\{t_i, m_i\}) \propto L_T(\underline{\theta})\pi(\underline{\theta})$$

Ahora, la prior conjunta esta dada por

$$\pi(\underline{\theta}) = f(\mu)f(a|\beta, c, p)f(\beta)f(c)f(p)$$

donde $f(\cdot)$ es la función de densidad marginal del parámetro y $f(\cdot|\cdot)$ es la función de densidad condicional.

Además,

$\mu \sim Ga(\gamma_\mu, \lambda_\mu)$ hace referencia a la secuencia del temblor *principal* o inicial y es independiente de los parámetros β, c, p que se refieren a los sismos subsecuentes (réplicas) y se distribuyen *Gamma*:

$\beta \sim Ga(\gamma_\beta, \lambda_\beta)$, describe la magnitud y es independiente de c, p que se refieren al tiempo.

$c \sim Ga(\gamma_c, \lambda_c)$

$p \sim U(0, 5)$

$a \sim Ga(\gamma_a, \lambda_a)$ y los hiperparametros (γ_a, λ_a) dependen de β, c, p .

Ahora bien, la prior $\pi(\underline{\theta})$ incorpora información de los datos antes de 1991 (con base en las mediciones del viejo sistema) y los datos obtenidos después de 1991 están incorporados en la verosimilitud $L_T(\underline{\theta})$ (con base en los datos obtenidos con el nuevo sistema de medición).

De lo anterior, se sabe que $\pi(\underline{\theta}|\{t_i, m_i\})$ no se puede obtener facilmente analíticamente ya que no es de forma cerrada. Existen métodos de simulación para calcular la distribución posterior conocidos como cadenas de Markov via Montecarlo: *MCMC*.

En particular, existe el método de *Metropolis*, el cual requiere únicamente que se pueda calcular el producto de la distribución inicial y la verosimilitud de los datos (que ya conocemos y fueron descritos).

$$\pi(\underline{\theta}) = \frac{1}{5} \lambda_\mu e^{-\lambda_\mu \mu} \lambda_\beta e^{-\lambda_\beta \beta} \lambda_c e^{-\lambda_c c} \lambda_a e^{-\lambda_a a} \frac{(\lambda_\mu \mu)^{\gamma_\mu - 1}}{\Gamma(\gamma_\mu)} \frac{(\lambda_\beta \beta)^{\gamma_\beta - 1}}{\Gamma(\gamma_\beta)} \frac{(\lambda_c c)^{\gamma_c - 1}}{\Gamma(\gamma_c)} \frac{(\lambda_a a)^{\gamma_a - 1}}{\Gamma(\gamma_a)}$$

Posteriormente, multiplicar la prior conjunta por la verosimilitud:

$$L_T(\underline{\theta}) = e^{\log(L_T(\underline{\theta}))}$$

donde,

$$\log(L_T(\underline{\theta})) = \log(\mu) + \sum_{i=1}^N \log(\mu + a + \sum_{t_j < t_i} \frac{e^{\beta(\mu_j - M_r)}}{(t_i - t_j + c)^p}) - \mu T - \frac{a}{p-1} \sum_{i=1}^N [e^{\beta(m_i - M_r)} (c^{-(p-1)} - (T - t_i + c)^{-(p-1)})]$$

Por lo que la distribución posterior es de la forma:

$$\pi(\theta|\{t_i, m_i\}) \propto e^{\log(\mu) + \sum_{i=1}^N \log(\mu+a + \sum_{t_j < t_i} \frac{e^{\beta(\mu_j - M_T)}}{(t_i - t_j + c)^p}) - \mu T - \frac{a}{p-1} \sum_{i=1}^N [e^{\beta(m_i - M_T)} (c^{-(p-1)} - (T - t_i + c)^{-(p-1)})]} \cdot \\ \lambda_\mu e^{-\lambda_\mu \mu} \lambda_\beta e^{-\lambda_\beta \beta} \lambda_c e^{-\lambda_c c} \lambda_a e^{-\lambda_a a} (\lambda_\mu \mu)^{\gamma_\mu - 1} (\lambda_\beta \beta)^{\gamma_\beta - 1} (\lambda_c c)^{\gamma_c - 1} (\lambda_a a)^{\gamma_a - 1}$$

Finalmente, el resultado obtenido con el algoritmo es una aproximación de la distribución posterior $\pi(\theta|\{t_i, m_i\})$ a partir de una muestra de θ .

7. Describan como conceptualizan el cálculo de la distribución predictiva para este modelo, y en qué diferiría respecto a los modelos que revisamos en clase.

En el artículo la conceptualización de la distribución es generada via el algoritmo Metropolis MCMC y del resultado se obtienen las medias para predecir los terremotos de magnitud mayor a 4.7. Esto difiere con lo visto en clase, pues usamos la verosimilitud y la distribución inicial para generar la distribución predictiva. Y se vio el uso del método gibbs que usa una muestra aleatoria de nuestra distribución, que es un caso particular del algoritmo Metropolis MCMC.

8. Para concluir, describan qué tipo de preguntas podrían ser relevantemente planteadas de manera que el modelo provea de una respuesta para ellas. (Consideren al menos dos preguntas).

- La probabilidad de que al menos un sismo ocurra en un par de horas o un intervalo de tiempo en Taiwan. Es relevante el lugar debido a las condiciones geográficas y placas tectónicas; además, los datos recopilados son de Taiwan, pues sería ilógico usarlos para predicciones en México por ejemplo.
- Probabilidad de que ocurra al menos un sismo dentro de los siguientes 15 minutos dado que acaba de ocurrir uno, es viable pues sólo sería la predicción de una nueva observación.
- La magnitud del siguiente sismo, es una predicción interesante que puede ser contestada con un rango con cota inferior como M_T . Una posible *mejora del modelo*, sería obtener cota inferior y superior pues no es lo mismo predecir un sismo de magnitud mayor a 4.7 (podría ser un terremoto de escal 9.8) que predecir un sismo con magnitud entre 4.7 y 5.