Inferencia sobre el modelo de recuento de partículas radiactivas (Caso 2)

Francisco A. Ocaña, Antonio Matilla y Mariano J. Valderrama

23 de febrero de 2009

Resumen

Este documento esboza la aplicación de la Teoría de la Probabilidad y la Inferencia Estadística en el problema de recuento de partículas radiactivas. En concreto, presenta la estimación de un modelo probabilístico que describe la emisión/recepción de partículas radiactivas a partir de recuentos estadísticamente independientes recabados en distintos instantes. Las explicaciones proporcionadas a lo largo de este documento son ilustradas en el fichero de Excel `radioactivt.xls`.

1. Modelo probabilístico

Consideremos el proceso de recuento de partículas radioactivas emitidas por una cierta fuente a lo largo del tiempo (medido en minutos, por ejemplo) y, posteriormente, recibidas por un cierto receptor. En este proceso, denotaremos con N_t al número de partículas recibidas hasta el instante t desde el inicio ($N_0 = 0$).

Desde un punto de vista matemático, se suele asumir que cada N_t es una variable aleatoria (v.a.) y, además, que la colección de vv.aa. $\{N_t\}_{t\geq 0}$, para el caso que nos ocupa, cumple las condiciones de los denominados *procesos estocásticos de Poisson* de parámetro λ, siendo dicho parámetro desconocido en la práctica. En concreto, $\{N_t\}$ cumple las siguientes propiedades:

- $N_t \rightsquigarrow \mathcal{P}\left(\lambda\, t\right), \ \ \forall t \geq 0$, donde $N_0 = 0$;
- el número de partículas en el intervalo de tiempo $(s, t]$, es decir, la diferencia $N_t - N_s$, cumple lo siguiente: $N_t - N_s \rightsquigarrow \mathcal{P}\left(\lambda(t-s)\right), \forall\, s < t$;
- los incrementos de $\{N_t\}$ son vv.aa. independientes.

Asimismo, puesto que

$$\mathrm{E}[N_t]/t = \lambda\,, \qquad \forall\, t > 0\,,$$

el parámetro λ se interpreta como el número medio de partículas recibidas por unidad de tiempo.

2. Inferencia sobre el modelo de recuento

En la práctica, como el parámetro λ del modelo de probabilidad presentado anteriormente es desconocido, éste ha de ser estimado a partir de recuentos obtenidos (datos) experimentalmente en el laboratorio. En concreto, supongamos que se realizan K recuentos (`cuentas`) independientes (poniendo a cero el dispositivo contador en cada uno de ellas en el laboratorio) de partículas radiactivas a distintos instantes $0 < t_1 < \ldots < t_K$, que supondremos ordenados de forma creciente. Este esquema de recuento experimental puede ser descrito matemáticamente a través de las K–variables aleatorias independientes dadas por $\{N_i = N_{t_i} : i = 1, \ldots, K\}$.

En este contexto, una forma de estimar el parámetro λ vendría dada a través de las vv.aa. independientes (`cpm` en el fichero `radiactivt.xls`)

$$\widehat{\lambda}_i \;=\; \frac{N_i}{t_i}\,, \qquad \forall\, i = 1, \ldots, K, \tag{1}$$

es decir, las tasas de partículas radiactivas por unidad de tiempo. En efecto, teniendo en cuenta que $N_i \rightsquigarrow \mathcal{P}\,(\lambda\, t_i)$, para cada i, pueden comprobarse las siguientes igualdades:

$$\begin{aligned} \mathrm{E}[\widehat{\lambda}_i] &= \frac{\mathrm{E}[N_i]}{t_i} = \lambda\,, && (2)\\ \mathrm{Var}[\widehat{\lambda}_i] &= \frac{\mathrm{Var}[N_i]}{t_i^2} = \frac{\lambda}{t_i} \quad \text{y} && \\ \sigma[\widehat{\lambda}_i] &= \sqrt{\frac{\lambda}{t_i}}\,, \qquad \forall\, i = 1, \ldots, K\,. && (3) \end{aligned}$$

Nótese que la Ecuación (2) nos confirma la idea de que $\widehat{\lambda}_i$ podría considerarse como un posible estimador del parámetro λ, siendo su error medido, teóricamente, a través de la Ecuación (3).

Una medida del decaimiento radiactivo es el *error estándar de cada estimador* $\widehat{\lambda}_i$, que para simplificar denotaremos con S_i, es decir,

$$S_i \;=\; \mathrm{se}(\widehat{\lambda}_i) \;=\; \hat{\sigma}(\widehat{\lambda}_i)\,. \tag{4}$$

Desde un punto de vista estadístico, cada S_i se interpreta como una medida del error de estimación de $\widehat{\lambda}_i$, como estimador que es de λ. Además, utilizando las Ecuaciones (1)–(4), podemos deducir una expresión para calcular S_i, a saber

$$S_i \;=\; \sqrt{\frac{\widehat{\lambda}_i}{t_i}} \;=\; \frac{\sqrt{N_i}}{t_i}\,, \qquad \forall\, i = 1, \ldots, K\,.$$

En la práctica, podemos ir calculando los estimadores $\{\widehat{\lambda}_i : i = 1, 2, \ldots\}$ junto con sus correspondientes errores $\{S_i : i = 1, 2, \ldots\}$ (ver hoja `Recuentos`, columnas `cpm` y `error`). De esta forma, disponemos de varias estimaciones alternativas para λ (una para cada recuento observado), junto con sus medidas de error asociado.

Si consideramos la aproximación de la distribución de Poisson a través de la Normal, podemos establecer que

$$\widehat{\lambda}_i \;\rightsquigarrow\; \frac{1}{t_i}\mathcal{P}\left(\lambda t_i\right) \;\approx\; \mathcal{N}\left(\lambda, S_i\right)\,.$$

De esta forma, se obtiene, para cada i, que

$$\mathrm{P}[\widehat{\lambda}_i - z_{\alpha/2} S_i \leq \lambda \leq \widehat{\lambda}_i + z_{\alpha/2} S_i] \approx 1 - \alpha\,,$$

siendo $\alpha \in (0, 1)$ y $\mathrm{P}[\mathcal{N}\left(0,1\right) > z_{\alpha/2}] = \alpha/2$. Así, podemos calcular intervalos de confianza (ICs) del parámetro λ para cada uno de los recuentos registrados en el laboratorio. Al representar gráficamente los extremos de dichos intervalos, para un cierto nivel de confianza $100(1-\alpha)$, podemos comprobar la enorme variabilidad entre ellos (ver gráfica en hoja `G.IC.Mi`) y, por tanto, la considerable imprecisión de estas estimaciones de λ.

Para resolver el problema anterior, observemos el comportamiento que presenta el error estándar S_i a lo largo del tiempo (ver gráfica en hoja `G.Si`). En base a este comportamiento, seleccionamos los recuentos con menor error (en nuestro caso, consideramos $S_i < 10$ en la hoja `Recuentos`), que como podemos comprobar corresponderán a los asociados a partir de un cierto índice i_o (los cinco últimos valores en nuestro caso). De esta forma, podríamos definir una nueva v.a. dada por

$$N = \sum_{i=i_o}^{K} N_i \rightsquigarrow \mathcal{P}\left(\lambda \sum_{i=i_o}^{K} t_i\right)\,, \tag{5}$$

sin más que aplicar la propiedad de reproductividad de la Poisson. Esta nueva v.a. nos conduce a un nuevo estimador de λ dado por

$$\hat{\lambda} \;=\; \frac{N}{\sum_{i=i_o}^{K} t_i} \;=\; \frac{\sum_{i=i_o}^{K} N_i}{\sum_{i=i_o}^{K} t_i}\,,$$

que cumple:

$$\begin{aligned} \mathrm{E}[\widehat{\lambda}] &= \lambda \quad \text{y} \\ S = \hat{\sigma}(\widehat{\lambda}) &= \sqrt{\frac{\widehat{\lambda}}{\sum_{i=i_o}^{K} t_i}} = \frac{\sqrt{N}}{\sum_{i=i_o}^{K} t_i} \end{aligned}$$

El lector podrá comprobar que esta nueva estimación de λ presenta un error más pequeño que las proporcionadas anteriormente (ver hoja `Recuentos`). De hecho, compare los ICs obtenidos para este nuevo estimador considerando la aproximación Normal de la Poisson (ver gráfica en hoja `G.IC.Mi`).

3. Validación del modelo estimado

El modelo de recuento estimado, dado con la estimación del parámetro λ, ofrece una descripción del comportamiento probabilístico del proceso de recuento. En ese sentido, vamos a estudiar el grado de adecuación de dicho modelo a los datos obtenidos en el laboratorio.

En primer lugar, podemos considerar el modelo estimado del número de partículas radiactivas que viene dado por

$$N_i \rightsquigarrow \mathcal{P}\left(\hat{\lambda} t_i\right) \approx \mathcal{N}\left(\hat{\lambda} t_i, \sqrt{\hat{\lambda} t_i}\right), \qquad \forall\, i = 1, \ldots, K\,.$$

De esta forma, podemos comparar los datos registrados para cada N_i con las bandas de confianza (extremos) obtenidas a partir de la distribución estimada de N_i (percentiles 97.5 % y 2.5 %, que dejan una masa de probabilidad de un 95 % entre ellos).

En segundo lugar, consideramos el modelo estimado de las tasas de partículas por unidad de tiempo.

$$\hat{\lambda}_i \rightsquigarrow \frac{1}{t_i}\mathcal{P}\left(\hat{\lambda} t_i\right) \approx \mathcal{N}\left(\hat{\lambda}, \sqrt{\frac{\hat{\lambda}}{t_i}}\right), \qquad \forall\, i = 1, \ldots, K\,.$$

De esta forma, podemos también comparar los datos registrados para cada $\hat{\lambda}_i$ con las bandas de confianza (extremos) obtenidas a partir de la distribución estimada de $\hat{\lambda}_i$ (percentiles 97.5 % y 2.5 %, que dejan una masa de probabilidad de un 95 % entre ellos).