UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS MATEMÁTICAS

FACULTAD DE CIENCIAS

Estimación de Probabilidades de Cruce de Nivel Desde una Perspectiva Bayesiana y de Grandes Desviaciones.

QUE PARA OBTENER EL GRADO ACADÉMICO DE

MAESTRO EN CIENCIAS

P R E S E N T A

César Almenara Martínez

DIRECTORA DE LA TESINA: Dra. Ana Meda Guardiola

MÉXICO, D.F. Marzo, 2012

Estimación de Probabilidades de Cruce de Nivel Desde una Perspectiva Bayesiana y de Grandes Desviaciones.

César Almenara Martínez.

Resumen

Este trabajo no pretende ser una revisión exhaustiva de las tres grandes teorías que aquí se encuentran: *Estadística Bayesiana no Paramétrica*, *Grandes Desviaciones* y *Teoría de Riesgo*. Se trata, más bien, de ejemplificar el uso de las mismas para la solución de un problema específico: estimar las probabilidades de cruce de nivel de un cierto proceso estocástico real-valuado que comienza en cero, en un contexto no paramétrico.

Tratamos con el problema de estimar la ley de una sucesión de variables aleatorias $\{X_k\}$, que toma valores en un espacio compacto Ω, mediante los *Procesos Dirichlet*. Que son procesos ampliamente estudiados y utilizados para la solución de problemas no paramétricos en la *Estadística Bayesiana*. Estos procesos resultan ser una familia conjugada de distribuciones apriori, cuya sucesión de distribuciones posteriores satisface el *Principio de Grandes Desviaciones*. Analizando las propiedades asintóticas de dicho proceso encontramos una expresión explícita para su función tasa y utilizamos estos resultados para resolver nuestro problema central aprovechándonos de algunos resultados procedentes de la *Teoría de riesgo*.

1 Introducción.

El estudio de probabilidades de cruce de nivel se ha convertido en un tema de gran relevancia en distintos campos, en particular en la *Teoría de Riesgo*. En lengua anglosajona, "*Theory of Risk*" es sinónimo de "*Non-Life Insurance Mathematics*" y abarca el estudio de la dinámica de portafolios de seguros homogéneos, es decir, contratos o pólizas cuyo riesgo es similar, como los seguros para autos, robo hogar, entreo otros, en donde surge el interés natural de preguntarse ¿cuál es la probabilidad de que una compañía de seguros (de daños) se aruine?, es decir, cuál es la probabilidad de que el proceso que modela el balance entre los ingresos y egresos de la compañía aseguradora sea menor que un cierto nivel positivo que puede representar, por ejemplo, su capital inicial. En dicho caso se dice que la compañía se ha aruinado o ha alcanzado la ruina. Este tipo de fenómenos son eventos de baja ocurrencia, es decir, eventos cuya probabilidad es pequeña, razón por la que se pueden modelar con la teoría de *Grandes Desviaciones*, teoría que estudia el comportamiento asintótico de sistemas estocásticos con función generadora de momentos. En este documento presentamos estimaciones asintóticas de probabilidades de cruce de nivel desde una perspectiva *Bayesiana* y de *Grandes Desviaciones*.

Existen dos propiedades deseables para una distribución apriori para problemas en la *Estadística Bayesiana* no paramétrica:

(I) El soporte de la distribución apriori debe ser grande con respecto a alguna topología adecuada en el espacio de distribuciones de probabilidad.

(II) Las distribuciones posteriores dada una muestra de observaciones a partir de la verdadera distribución de probabilidades deben ser manipulables analíticamente.

Estas propiedades son antagónicas en el sentido de que una se puede obtener a expensas de la otra. *Thomas S. Ferguson* [Fer73] presenta una clase de distribuciónes apriori, llamada **Procesos Dirichlet**, basada en el sentido de (I), para la cual (II) se cumple y además tiene la ventaja de que para muchos problemas de la estadística no paramétrica es aplicable y proporciona resultados satisfactorios [como se expone en las secciones 2 y 3].

Estos procesos tienen la propiedad de ser consistentes con la teoría de Grandes Desviaciones, de modo que si las distribuciones apriori son un Proceso Dirichlet, entonces la sucesión de distribuciones posteriores satisface un Principio de Grandes Desviaciones y podemos encontrar una expresión explícita para la función tasa [Sección 4]. Aprovechando la relación existente entre los Porcesos Dirichlet y las Grandes Desviaciones, presentamos aproximaciones asintóticas para las probabilidades de cruce de nivel en un contexto Bayesiano [Sección 5].

La principal referencia de este trabajo es el artículo de *C. Macci y L. Petrella* [MP06], en el que se basa la exposición de la sección 5. La parte de los *Procesos Dirichlet* presentada en las secciones 2 y 3, está fundamentada en [Fer73], mientras que la sección 4, se basa en [GO00].

2 Distribución Dirichlet

El contenido de esta sección es un campo ampliamente estudiado, sin embargo la definición de la *Distribución Dirichlet* establecida por *T. S. Ferguson* en [Fer73] es ligeramente más general que la usual y de suma importancia para este documento, ya que parte del desarrollo del mismo se basa en esta definición y en las propiedades establecidas en esta sección. Para mayor referencia sobre el tema, el lector interesado puede revisar por ejemplo el libro de *S.S Wilks* [Wil62].

La *Distribución Dirichlet* es conocida por los *Bayesianos* como la conjugada apriori para los parámetros de una distribución multinomial [Goo65].

Denotemos por $Gama(\cdot|a,\beta)$ a la distribución gama con parámetro de forma $a \geq 0$ y parámetro de escala $\beta > 0$. Para $a = 0$ esta distribución es degenerada en el origen, y para $a > 0$ tiene densidad con respecto a la medida de *Lebesgue* en la línea real dada por

$$f(z|a,\beta) = \frac{1}{\Gamma(a)\beta^a} z^{a-1}\beta e^{-z/\beta}\mathbb{I}_{\{0,\infty\}}(z), \tag{1}$$

La pequeña generalización que *Ferguson* presenta en su definición permite que algunas de las variables involucradas sean degeneradas en cero, como podemos ver a continuación.

Definición 1. *Sean $Z_1, Z_2, \ldots, Z_k$ variables aleatorias idénticamente distribuidas, con distribución $Gama(\cdot|a_j,1)$ donde $a_j \geq 0$ para todo j, y $a_j > 0$ para alguna j, $j = 1,2,\ldots,k$. La* Distribución Dirichlet *con parámetros $(a_1,\ldots,a_k)$, denotada por $\mathscr{D}(a_1,\ldots,a_k)$, está definida como la distribución conjunta de $(Y_1,\ldots,Y_k)$ con*

$$Y_j = \frac{Z_j}{\sum_{i=1}^k Z_i}, \quad j = 1,2,\ldots,k. \tag{2}$$

Esta distribución siempre es singular con respecto a la medida de *Lebesgue* en un espacio k-dimensional, pues $Y_1 + Y_2 + \cdots + Y_k = 1$. Adicionalmente, si para algún j con $j = 1,\ldots,k$, $a_j = 0$, entonces la correspondiente Y_j es degenerada en el origen. No obstante, si $a_j > 0$ para todo j, la distribución $(k-1)$-dimensional del vector aleatorio $(Y_1,\ldots,Y_{k-1})$ es absolutamente continua con densidad

$$f(y_1,\ldots,y_{k-1}|a_1,\ldots,a_k) = \frac{\Gamma(a_1+\cdots+a_k)}{\Gamma(a_1)\cdots\Gamma(a_k)}\left(\prod_{j=1}^{k-1} y_j^{a_k-1}\right)\left(1-\sum_{j=1}^{k-1} y_j\right)^{a_k-1}\mathbb{I}_S(y_1,\ldots,y_{k-1}), \tag{3}$$

donde $S = \{(y_1,\ldots,y_{k-1})|\ y_j \geq 0,\ \sum_{j=1}^{k-1} y_j \leq 1\}$.

Para $k = 2$, (3) se reduce a la distribución beta con parámetros a_1 y a_2 denotada por $Beta(\cdot|a_1,a_2)$.

A continuación enunciaremos algunas propiedades de la *Distribución Dirichlet*, poniendo especial énfasis en la primera de ellas.

1. Si $(Y_1,\ldots,Y_k) \sim \mathscr{D}(a_1,\ldots,a_k)$ y $r_1,\ldots,r_l$ son enteros tales que $0 < r_1 < \cdots < r_l = k$, entonces

$$\left(\sum_{i=1}^{r_1} Y_i, \sum_{i=r_1+1}^{r_2} Y_i, \ldots, \sum_{i=r_{l-1}+1}^{r_l} Y_i\right) \sim \mathscr{D}\left(\sum_{i=1}^{r_1} a_i, \sum_{i=r_1+1}^{r_2} a_i, \ldots, \sum_{i=r_{l-1}+1}^{r_l} a_i\right).$$

   Este hecho se sigue directamente de la definición de la *Distribución Dirichlet* y la propiedad de aditividad para variables aleatorias gama independientes con mismo parámetro de escala. En particular bajo estas condiciones, la distribución marginal de cada Y_j es $Beta(Y_j|a_j,(\sum_{i=1}^k a_i) - a_j)$.

2. Si $(Y_1, \ldots, Y_k) \sim \mathscr{D}(a_1, \ldots, a_k)$ entonces

$$\begin{aligned}
\mathbb{E}[Y_i] &= \frac{a_i}{a} \\
\mathbb{E}[Y_i^2] &= \frac{a_i(a_i+1)}{a(a+1)} \\
\mathbb{E}[Y_iY_j] &= \frac{a_ia_j}{a(a+1)}, \quad \text{para} \quad i \neq j,
\end{aligned}$$

donde $a = \sum_{i=1}^{k} a_i$.

3. La siguiente propiedad de la *Distribución Dirichlet* está apliamente relacionada con el *Teorema de Bayes*. Si la distribución apriori del vector aleatorio $(Y_1, \ldots, Y_k)$ es $\mathscr{D}(a_1, \ldots, a_k)$ y si $\mathbb{P}\{X = j | Y_1, \ldots, Y_k\} = Y_j$ casi seguramente bajo $\mathbb{P}$ para $j = 1, \ldots, k$, entonces la distribución posterior del vector $(Y_1, \ldots, Y_k)$ dado $X = j$ es $\mathscr{D}(a_1^{(j)}, \ldots, a_k^{(j)})$ donde

$$a_i^{(j)} = \begin{cases} a_j & \text{si } i \neq j \\ a_j + 1 & \text{si } i = j \end{cases}$$

Denotemos a la función de distribución *Dirichlet* de parámetros $a_1, \ldots, a_k$, como $D(y_1, \ldots, y_k | a_1, \ldots, a_k)$. Entonces la igualdad

$$\mathbb{P}\{X = j, Y_1 \leq z_1, \ldots, Y_k \leq z_k\} = \mathbb{P}\{X = j\}\mathbb{P}\{Y_1 \leq z_1, \ldots, Y_k \leq z_k | X = j\}, \tag{4}$$

puede escribirse en términos de la función *Dirichlet* (utilizando las propiedades 2 y 3) como sigue

$$\int_0^{z_1} \cdots \int_0^{z_k} y_j \mathrm{d}D(y_1, \ldots, y_k | a_1, \ldots, a_k) = \frac{a_j}{a} D(z_1, \ldots, z_k | a_1^{(j)}, \ldots, a_k^{(j)}). \tag{5}$$

Es importante hacer notar que (5) sigue valiendo aún cuando $a_j = 0$.

3 El Proceso Dirichlet.

El *Proceso Dirichlet* introducido por *Ferguson* en 1973 [Fer73] posteriormente tratado por *Blackwell* y *MacQueen* [BM73] en ese mismo año es la base de la variedad más amplia de modelos utilizados en la estadística *Bayesiana* no paramétrica, principalmente debido a la actual disponibilidad de técnicas computacionales eficientes, pero también, por la gran diversidad en las aplicaciones de estas técnicas no parámetricas en distintas disciplinas como las finanzas, genética, medicina entre muchas otras.

Definamos el espacio $\mathcal{M}$, como el conjunto de todas las medidas μ en $\mathbb{R}^+$ tales que $\mu(0) = 0$ y $\mu(t) < \infty$ para toda $t \in \mathbb{R}^+$ y sea $\mathcal{B}(\mathcal{M})$ la σ-álgebra de Borel en $\mathcal{M}$.

Definición 2. *Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Si $P : (\Omega, \mathcal{F}, \mathbb{P}) \longmapsto (\mathcal{M}, \mathcal{B}(\mathcal{M}))$ es un mapeo medible, entonces decimos que P es una medida aleatoria. Si además P es una medida de probabilidad, diremos que P es una medida de probabilidad aleatoria.*

De aquí en adelante, salvo cualquier otra especificación, sea $\mathcal{X}$ un conjunto cualquiera no vacío y $\mathcal{A}$ una σ-álgebra de subconjuntos de $\mathcal{X}$.

Definición 3. *Sea α una medida no nula y finita en $(\mathcal{X}, \mathcal{A})$. Diremos que una medida de probailidad aleatoria P, es un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α, si para cada $k = 1, 2, \dots$ y para toda partición medible*[1] *$(B_1, \dots, B_k)$ de $\mathcal{X}$, la distribución de $(P(B_1), \dots, P(B_k))$ es Dirichlet, $\mathscr{D}(\alpha(B_1), \dots, \alpha(B_k))$.*

Las condiciones de medibilidad y consistencia de este proceso se encuentran demostradas y estudiadas en [Fer73]. Las siguientes tres proposiciones muestran la relación existente entre las propiedades del proceso y las propiedades de su parámetro.

Proposición 1. *Sean P un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α y $A \in \mathcal{A}$. Si $\alpha(A) = 0$, entonces $P(A) = 0$ casi seguramente. Y si $\alpha(A) > 0$, entonces $P(A) > 0$ con probabilidad 1. Aún más*

$$\mathbb{E}[P(A)] = \frac{\alpha(A)}{\alpha(\mathcal{X})}.$$

Demostración.

Tomemos la partición medible (A, A^c), entonces $(P(A), P(A^c))$ se distribuye Dirichlet, $\mathscr{D}(\alpha(A), \alpha(A^c))$. Como consecuencia inmediata de la propiedad (3) de la distribución Dirichlet, $P(A)$ tiene distribución beta de parámetros $\alpha(A)$ y $\alpha(A^c)$, de modo que si $\alpha(A) = 0$ casi seguramente, entonces $P(A)$ es degenerada en cero y por lo tanto $P(A) = 0$. Analogamente si $\alpha(A) > 0$ casi seguramente, tenemos que $P(A) > 0$ con probabilidad 1. Dado que $P(A) \sim Beta(P(A)|\alpha(A), \alpha(A^c))$, podemos concluir además que

$$\mathbb{E}[P(A)] = \frac{\alpha(A)}{\alpha(A) + \alpha(A^c)} = \frac{\alpha(A)}{\alpha(\mathcal{X})}$$

$\diamond$

Esta proposición puede ser un poco engañosa ya que aparentemente α y P tienen los mismos conjuntos nulos, es decir que son mutuamente absolutamente continuas. Sin embargo, esto es falso ya que como se muestra en [Fer73] y se explica brevemente en la siguiente sección, P es esencialmente una distribución discreta. Entonces α y P podrían ser mutuamente singulares. Para evitar este tipo de problemas y quitar la limitante de asignar probabilidad uno al espacio de distribuciones discretas, se han desarrollado generalizaciones como la presentada por *Antoniak* [Ant74]. Sin embargo, este tipo de modelos van más allá de las necesidades de este documento.

Proposición 2. *Sea P un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α. Si α es σ-aditiva, entonces P también lo es, en el sentido de que para una sucesión fija decreciente de conjuntos medibles $A_n \searrow \varnothing$, entonces $P(A_n) \to 0$ con probabilidad* 1.

[1] $(B_1, \dots, B_k)$ es una partición medible de $\mathcal{X}$ si $B_i \in \mathcal{A}$ para todo i, $B_i \cap B_j = \varnothing$ para $i \neq j$ y además $\bigcup_{j=1}^{k} B_j = \mathcal{X}$.

Demostración.

Como $A_n \searrow \varnothing$ y α es σ-aditiva, $A_n \supset A_{n+1}$ $\forall n \geq 1$, $\bigcap_{n\in\mathbb{N}} A_n = \varnothing$ y entonces $\alpha(A_n) \to 0$ cuando n tiende a infinito. De modo que existe una sub-sucesión $\{n_j\}$ tal que $\sum_{j=1}^{\infty} \alpha(A_{n_j}) < \infty$, por lo tanto para un $\epsilon > 0$ la desigualdad de *Markov* aplicada a $P(A_{n_j})$, junto con el resultado de la propiedad 1, nos lleva a concluir que

$$\sum_{j=1}^{\infty} \mathbb{P}\{P(A_{n_j}) > \epsilon\} \leq \sum_{j=1}^{\infty} \frac{\mathbb{E}[P(A_{n_j})]}{\epsilon} = \frac{1}{\epsilon} \sum_{j=1}^{\infty} \frac{\alpha(A_{n_j})}{\alpha(\mathcal{X})} < \infty.$$

Al aplicar la primera parte del lema de *Borel-Cantelli* tenemos que

$$\mathbb{P}\{\limsup_{j\to\infty} P(A_{n_j}) > \epsilon\} = 0,$$

lo que prueba que $P(A_{n_j}) \to 0$, con probabilidad uno. Por último sólo basta recalcar que $P(A_n) > P(A_{n+1})$ para toda $n \geq 1$ de manera casi segura, y entonces $P(A_1) > P(A_2) > \cdots$ casi seguramente, con lo que se concluye el resultado.

$\diamond$

Proposición 3. *Sea P un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α y sea Q una medida de probabilidad fija en $(\mathcal{X}, \mathcal{A})$ tal que $Q \ll \alpha$. Entonces para cualquier entero positivo m y cualesquiera conjuntos medibles $A_1, \ldots, A_m$, se cumple que*

$$\mathbb{P}\{|P(A_i) - Q(A_i)| < \epsilon \quad para \quad i = 1, 2, \ldots, m\} > 0 \quad \forall \epsilon > 0.$$

Demostración.

Tomemos m y $A_1, \ldots, A_m$ como en el enunciado y definamos $B_{\nu_1,\ldots,\nu_m}$ para cada $\nu_j = 0$ o 1 como sigue:

$$B_{\nu_1,\ldots,\nu_m} := \bigcap_{j=1}^{m} A_j^{\nu_j}, \tag{6}$$

donde A_j^1 es interpretado como A_j y A_j^0 como A_j^c. Es decir que (6) forma los 2^m conjuntos obtenidos al tomar intersecciones de los A_i y sus complementos. Claramente cada uno de estos conjuntos $B_{\nu_1,\ldots,\nu_m}$ pertenece a $\mathcal{A}$, además los 2^m conjuntos resultantes son disjuntos. De modo que al unirlos obtenemos $\mathcal{X}$. Por lo tanto $\{B_{\nu_1,\ldots,\nu_m}\}$ forma una partición medible de $\mathcal{X}$.

Ahora, notemos que

$$\begin{aligned} &\mathbb{P}\{|P(A_i) - Q(A_i)| < \epsilon \quad para \quad i = 1, 2, \ldots, m\} \geq \\ &\qquad \mathbb{P}\{\textstyle\sum_{\nu_1,\ldots,\nu_m|\nu_i=1} |P(B_{\nu_1,\ldots,\nu_m}) - Q(B_{\nu_1,\ldots,\nu_m})| < \epsilon \quad para \quad i = 1, \ldots, m\}, \end{aligned} \tag{7}$$

ya que

$$A_i = \bigcup_{\nu_1,\ldots,\nu_m|\nu_i=1} B_{\nu_1,\ldots,\nu_m} \quad \forall \quad i = 1, \ldots, m,$$

entonces

$$P(A_i) = \sum_{\nu_1,\ldots,\nu_m|\nu_i=1} P(B_{\nu_1,\ldots,\nu_m}) \quad \text{y} \quad Q(A_i) = \sum_{\nu_1,\ldots,\nu_m|\nu_i=1} Q(B_{\nu_1,\ldots,\nu_m}),$$

lo que implica la validez de (7) al utilizar la desigualdad del triángulo y la monotonía de $\mathbb{P}$. Por lo tanto, es suficiente demostrar que

$$\mathbb{P}\{|P(B_{\nu_1,\ldots,\nu_m}) - Q(B_{\nu_1,\ldots,\nu_m})| < \frac{\epsilon}{2^m} \quad \forall \quad (\nu_1,\ldots,\nu_m)\} > 0,$$

pues al sumar sobre los 2^m índices distintos $(\nu_1,\ldots,\nu_m)$, se obtiene que la parte derecha de la ecuación (7) es estrictamente positiva y de este modo habremos probado el resultado propuesto.

Con esto en mente, como $Q \ll \alpha$, si $\alpha(B_{\nu_1,\ldots,\nu_m}) = 0$, entonces $Q(B_{\nu_1,\ldots,\nu_m}) = 0$. Utilizando la primera parte de la proposición 1, $P(B_{\nu_1,\ldots,\nu_m}) = 0$ con probabilidad uno, lo que nos permite concluir que $|P(B_{\nu_1,\ldots,\nu_m}) - Q(B_{\nu_1,\ldots,\nu_m})| = 0$ casi seguramente; mientras que para aquellos índices $(\nu_1,\ldots,\nu_m)$ para los que $\alpha(B_{\nu_1,\ldots,\nu_m}) > 0$ la distribución del correspondiente $P(B_{\nu_1,\ldots,\nu_m})$ otorga peso positivo a todos los conjuntos abiertos en el conjunto $\sum_{\nu_1,\ldots,\nu_m|\alpha(B_{\nu_1,\ldots,\nu_m})>0} P(B_{\nu_1,\ldots,\nu_m}) = 1$, con lo que se concluye la prueba.

◇

Definición 4. *Sea P una medida de probabilidad aleatoria en $(\mathcal{X},\mathcal{A})$. Decimos que $X_1,\ldots,X_n$ es una muestra de tamaño n tomada de P, si para cualquier $m = 1,2,\ldots$ y cualesquiera conjuntos medibles $A_1,\ldots,A_m$, $C_1,\ldots,C_n$,*

$$\mathbb{P}\{X_1 \in C_1,\ldots,X_n \in C_n|P(A_1),\ldots,P(A_m),P(C_1),\ldots,P(C_n)\} = \prod_{j=1}^{n} P(C_j) \quad c.s. \tag{8}$$

Estrictamente hablando, $X_1,\ldots,X_n$ es una muestra de tamaño n tomada de P, si dados $P(C_1),\ldots,P(C_n)$, los eventos $\{X_1 \in C_1\},\ldots,\{X_n \in C_n\}$ son independientes del resto del proceso y son independientes entre si, con $\mathbb{P}\{X_j \in C_j|P(C_1),\ldots,P(C_n)\} = P(C_j)$ casi seguramente para $j = 1,\ldots,n$.

Esta definición determina la distribución conjunta de $X_1,\ldots,X_n, P(A_1),\ldots,P(A_m)$, una vez que la distribución del proceso está dada, ya que

$$\mathbb{P}\{X_1 \in C_1,\ldots,X_n \in C_n, P(A_1) \le y_1,\ldots,P(A_m) \le y_m\}, \tag{9}$$

puede obtenerse al integrar la ecuación (8) con respecto a la distribución conjunta de $P(A_1),\ldots,P(A_m)$, $P(C_1),\ldots,P(C_n)$ sobre el conjunto $[0,y_1] \times \cdots \times [0,y_m] \times [0,1] \times \cdots \times [0,1]$.

Proposición 4. *Sea P un* Proceso Dirichlet *en $(\mathcal{X},\mathcal{A})$ con parámetro α y sea X una muestra de tamaño 1 tomada de P. Entonces para $A \in \mathcal{A}$,*

$$\mathbb{P}\{X \in A\} = \frac{\alpha(A)}{\alpha(\mathcal{X})}.$$

Demostración.

$$\begin{aligned}\mathbb{P}\{X \in A\} &= \mathbb{E}[\mathbb{E}[\mathbb{I}_{\{X\in A\}}|P(A)]]\\ &= \mathbb{E}[\mathbb{P}\{X \in A|P(A)\}]\\ &= \mathbb{E}[P(A)]\\ &= \frac{\alpha(A)}{\alpha(\mathcal{X})},\end{aligned}$$

donde la tercer igualdad es válida ya que X es una muestra de tamaño 1 tomada de P, y por lo tanto $\mathbb{P}\{X \in A|P(A)\} = P(A)$ casi seguramente, mientras que la última igualdad se sigue de la propiedad 1.

◇

Proposición 5. *Sea P un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α y sea X una muestra de tamaño 1 tomada de P. Además sean $(B_1, \ldots, B_k)$ una partición medible de $\mathcal{X}$ y $A \in \mathcal{A}$. Entonces*

$$\mathbb{P}\{X \in A, P(B_1) \leq y_1, \ldots, P(B_k) \leq y_k\} = \sum_{j=1}^{k} \frac{\alpha(B_j \cap A)}{\alpha(\mathcal{X})} D(y_1, \ldots, y_k|\alpha_1^{(j)}, \ldots, \alpha_k^{(j)}), \quad (10)$$

donde $D(y_1, \ldots, y_k|\alpha_1, \ldots, \alpha_k)$ es la función de distribución de la Distribución Dirichlet *$\mathscr{D}(\alpha_1, \ldots, \alpha_k)$ y*

$$\alpha_i^{(j)} = \begin{cases} \alpha(B_i) & si\ i \neq j \\ \alpha(B_j) + 1 & si\ i = j \end{cases}$$

Demostración.

Definamos primero los siguientes conjuntos, $B_{j,1} = B_j \cap A$ y $B_{j,0} = B_j \cap A^c$ para $j = 1, \ldots, k$. Sea $Y_{j,\nu} = P(B_{j,\nu})$ para $j = 1, \ldots, k$ y $\nu = 0$ o 1. Entonces, aplicando (8) tenemos que

$$\begin{aligned}\mathbb{P}\{X \in A|Y_{j,\nu} \text{ con } j = 1, \ldots, k \quad \text{y} \quad \nu = 0, 1\} = P(A) &= P\Big(\bigcup_{j=1}^{k} B_j \cap A\Big)\\ &= \sum_{j=1}^{k} P(B_j \cap A) = \sum_{j=1}^{k} P(B_{j,1}) \qquad (11)\\ &= \sum_{j=1}^{k} Y_{j,1}, \qquad c.s.\end{aligned}$$

De modo que para $y_{j,\nu} \in [0, 1]$ arbitrarios con $j = 1, \ldots, k$ y $\nu = 0, 1$

$$\mathbb{P}\{X \in A, Y_{j,\nu} \leq y_{j,\nu} \quad \text{para} \quad j = 1, \ldots, k \quad \text{y} \quad \nu = 0, 1\},$$

se encuentra integrando (11) con respecto a la distribución de $Y_{j,\nu}$ sobre el conjunto $\{Y_{j,\nu} \leq y_{j,\nu}, j = 1, \ldots, k \text{ y } \nu = 0, 1\}$. Utilizando las ecuaciones (4) y (5) obtenemos que el resultado de esta integral es

$$\frac{\alpha(A)}{\alpha(\mathcal{X})} D(\mathbf{y}|\alpha^{(\mathbf{j})}) = \sum_{j=1}^{k} \frac{\alpha(B_{j,1})}{\alpha(\mathcal{X})} D(\mathbf{y}|\alpha^{(\mathbf{j})}),$$

en donde $\mathbf{y} = (y_{1,0}, \ldots, y_{k,0}, y_{1,1}, \ldots, y_{k,1})$, $\alpha^{(\mathbf{j})} = (\alpha_{1,0}^{(j)}, \ldots, \alpha_{k,0}^{(j)}, \alpha_{1,1}^{(j)}, \ldots, \alpha_{k,1}^{(j)})$ y

$$\alpha_{i,\nu}^{(j)} = \begin{cases} \alpha(B_{i,\nu}) & \text{si } i \neq j \\ \alpha(B_{j,\nu}) + 1 & \text{si } i = j \end{cases}$$

Por último notemos que $P(B_j) = Y_{j,0} + Y_{j,1}$ casi seguramente, entonces al utilizar la propiedad (1) de la *Distribución Dirichlet* y el hecho de que el proceso de encontrar distribuciones marginales de variables aleatorias es lineal, podemos afirmar que la ecuación (10) es válida y concluir la prueba.

◇

La proposición anterior nos da las herramientas necesarias para calcular la distribución conjunta de un *Proceso Dirichlet* P, dada una muestra $X_1, \ldots, X_n$ tomada de P.

Para $x \in \mathcal{X}$, sea δ_x la medida en $(\mathcal{X}, \mathcal{A})$ que da masa uno al punto x (delta de *Dirac*).

$$\delta_x = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

Teorema 1. *Sea P un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α, y sea $X_1, \ldots, X_n$ una muestra de tamaño n tomada de P. Entonces la distribución condicional de P dada la muestra $X_1, \ldots, X_n$, es la de un* Proceso Dirichlet *con parámetro $\alpha + \sum_{i=1}^{n} \delta_{X_i}$.*

Demostración.

Para verificar este resultado es suficiente demostrar el teorema para $n = 1$, ya que para $n > 1$ el resultado se sigue inmediatamente por inducción bajo la repetida aplicación del caso $n = 1$. Tomemos $(B_1, \ldots, B_k)$ una partición medible de $\mathcal{X}$ y sea $A \in \mathcal{A}$. El objetivo es demostrar que la distribución condicional de $P(B_1), \ldots, P(B_k)$ dado X, una muestra de tamaño 1 tomada de P, tiene función de distribución igual a

$$D(y_1, \ldots, y_k | \alpha(B_1) + \delta_x(B_1), \ldots, \alpha(B_k) + \delta_x(B_k)). \tag{12}$$

Para lograr esto, integramos (12) con respecto a la distribución marginal de X sobre A.

$$\begin{aligned} &\int_A D(y_1, \ldots, y_k | \alpha(B_1) + \delta_x(B_1), \ldots, \alpha(B_k) + \delta_x(B_k)) \frac{\mathrm{d}\alpha(x)}{\alpha(\mathcal{X})} \\ &\qquad = \sum_{j=1}^{k} \int_{B_j \cap A} D(y_1, \ldots, y_k | \alpha_1^{(j)}, \ldots, \alpha_k^{(j)}) \frac{\mathrm{d}\alpha(x)}{\alpha(\mathcal{X})} \\ &\qquad = \sum_{j=1}^{k} \frac{\alpha(B_j \cap A)}{\alpha(\mathcal{X})} D(y_1, \ldots, y_k | \alpha_1^{(j)}, \ldots, \alpha_k^{(j)}), \end{aligned}$$

donde la primer igualdad hace uso de los siguientes hechos

$$\bigcup_{j=1}^{k} B_j \cap A = A, \qquad \alpha_i^{(j)} = \begin{cases} \alpha(B_i) & \text{si } i \neq j \\ \alpha(B_j)+1 & \text{si } i = j \end{cases}, \qquad i,j = 1,\ldots,k$$

$$\delta_x(B_i) = \begin{cases} 1 & \text{si } x \in B_i \\ 0 & \text{si } x \notin B_i \end{cases},$$

junto con la linealidad de la integral, mientras que la segunda igualdad se sigue bajo simple integración.

Por lo tanto la distribución condicional de $P(B_1),\ldots,P(B_k)$ dado X tiene función de distribución igual a (12).

$\diamond$

3.1 Una Definición Alternativa del Proceso Dirichlet.

A continuación se da la difinición de una medida de probabilidad que resulta ser un *Proceso Dirichlet* en $(\mathcal{X},\mathcal{A})$ con parámetro α y con probabilidad 1 es una medida de probabilidad discreta en $(\mathcal{X},\mathcal{A})$.

La idea básica es que dada la definición de la *Distribución Dirichlet* expuesta en (2) como la distribución conjunta de un conjunto de variables aleatorias con distribución gama independientes entre si y divididas entre la suma de las mismas, podemos pensar en definir el *Proceso Dirichlet* como un proceso gama[2] con incrementos independientes divididos por su suma. Utilizando una representación de un proceso con incrementos independientes como una suma de un número numerable de saltos de altura aleatoria en un número numerable de puntos aleatorios y dividir por el total de alturas de los saltos, obteniendo así una medida de probabilidad discreta, que debe distribuirse como un *Proceso Dirichlet* (ver [FK72] para mayores detalles sobre el como se obtiene la medida de probabilidad discreta).

La distribución gama, $Gama(X|\alpha,1)$ con $\alpha > 0$, tiene función característica (ver *Gnedenko y Kolmogorov* [BK54] pp 86-87)

$$\varphi(t) = (1-it)^{-\alpha} = \exp\Big\{\int_0^\infty (e^{itx}-1)\mathrm{d}N(x)\Big\}, \tag{13}$$

donde

$$N(x) = -\alpha \int_x^\infty e^{-y}y^{-1}\mathrm{d}y, \qquad 0 < x < \infty. \tag{14}$$

Definimos la distribución de las variables aleatorias $J_1, J_2, \ldots$ (los saltos) como sigue

$$\mathbb{P}\{J_1 \leq x_1\} = e^{N(x_1)}, \qquad \text{para} \quad x_1 > 0, \tag{15}$$

y para $j = 2,3,\ldots$

$$\mathbb{P}\{J_j \leq x_j | J_{j-1} = x_{j-1},\ldots,J_1 = x_1\} = \exp\{N(x_j) - N(x_{j-1})\}, \quad \text{para} \quad 0 < x_j < x_{j-1}. \tag{16}$$

[2]Un proceso gama es un proceso estocástico con incrementos independientes de distribución gama

De modo que la distribución de J_1 es $\exp\{N(x_1)\}$ y para $j = 2,3,\ldots$ la distribución de J_j dados $J_{j-1},\ldots,J_1$, es la misma que la distribución de J_1 truncada por arriba en J_{j-1}.

El siguiente teorema está basado en el teorema principal de [FK72].

Teorema 2. *Sea $G(t)$ una función de distribución en $[0,1]$. Definamos*

$$Z_t = \sum_{j=1}^{\infty} J_j \mathbb{I}_{[0,G(t))}(U_j), \tag{17}$$

donde

(i) La distribución de $J_1, J_2, \ldots$ está dada como en (15) y (16).

(ii) $U_1, U_2, \ldots$ son variables aleatorias independientes e idénticamente distribuidas, con distribución uniforme en $[0,1]$ e independientes de $J_1, J_2, \ldots$. Entonces con probabilidad uno, Z_t converge para toda $t \in [0,1]$ y es un proceso gama con incrementos independientes, con $Z_t \sim Gama(\cdot|\alpha G(t), 1)$.

La demostración del teorema anterior será referida a [FK72], donde se establece la generalización del resultado mencionado arriba y cuya prueba se basa en el desarrollo de cuatro lemas demostrados a lo largo de todo el artículo.

Al analizar el resultado del teorema anterior, nos damos cuenta de que en particular $Z_1 = \sum_{j=1}^{\infty} J_j$ converge con probabilidad uno y además $Z_1 \sim Gama(Z|\alpha, 1)$. De tal modo que si definimos

$$P_j := \frac{J_j}{Z_1} \tag{18}$$

entonces $P_j \geq 0$ y $\sum_{j=1}^{\infty} P_j = 1$ casi seguramente.

Basados en lo anterior, definimos el *Proceso Dirichlet* como sigue. Tomemos de nuevo $(\mathcal{X}, \mathcal{A})$ un espacio medible, $\alpha(\cdot)$ una medida finita no nula en $\mathcal{A}$ y sea $V_1, V_2, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con valores en $\mathcal{X}$ y con medida de probabilidad Q, donde $Q(A) := \frac{\alpha(A)}{\alpha(X)}$.

Identifiquemos el α de las fórmulas (13) y (14) con $\alpha(\mathcal{X})$, y definamos la medida de probabilidad aleatoria P, en $(\mathcal{X}, \mathcal{A})$ como

$$P(A) := \sum_{j=1}^{\infty} P_j \delta_{V_j}(A). \tag{19}$$

Teorema 3. *La medida de probabilidad aleatoria definida en (19) es un* Proceso Dirichlet *en $(\mathcal{X}, \mathcal{A})$ con parámetro α.*

Demostración.

Sea $(B_1, \ldots, B_k)$ una partición medible de $\mathcal{X}$. Entonces dada la definición de la ecuación (19) tenemos que

$$(P(B_1), \ldots, P(B_k)) = \frac{1}{Z_1} \sum_{j=1}^{\infty} J_j (\delta_{V_j}(B_1), \ldots, \delta_{V_j}(B_k)).$$

Como por hipótesis $V_1, V_2, \ldots$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, entonces

$$M_j := (\delta_{V_j}(B_1), \ldots, \delta_{V_j}(B_k)),$$

son vectores aleatorios independientes e idénticamente distribuidos con distribución mulltinomial y vector de probabilidades $(Q(B_1), \ldots, Q(B_k))$. Por lo tanto, la distribución de $\sum_{j=1}^{\infty} J_j M_j$ debe tener, según el resultado del teorema 2, la misma distribución que

$$\left(Z_{\frac{1}{k}}, Z_{\frac{2}{k}} - Z_{\frac{1}{k}}, \ldots, Z_1 - Z_{\frac{k-1}{k}}\right),$$

donde Z_t es el proceso gama definido por (17), tomando $G(t)$ como

$$G\left(\frac{j}{k}\right) - G\left(\frac{j-1}{k}\right) = Q(B_j), \qquad j = 1, \ldots, k.$$

Lo que implica que $\sum_{j=1}^{\infty} J_j \delta_{V_j}(B_i)$ son variables aleatorias independientes con distribución $Gama(\cdot|\alpha(B_i), 1)$, para $i = 1, \ldots, k$ (resultado del teorema 2). Por otro lado, como Z_1 es la suma de estas variables aleatorias gama, entonces $(P(B_1), \ldots, P(B_k)) \sim \mathscr{D}(\alpha(B_1), \ldots, \alpha(B_k))$ según nuestra definición de la *Distribución Dirichlet*. Por lo tanto, P satisface la definición de un *Proceso Dirichlet*.

◇

Con la finalidad de enfatizar la relación existente entre α y la medida de probabilidad aleatoria P, exponemos el siguiente resultado.

Teorema 4. *Sea P el* Proceso Dirichlet *definido en (19), y sea Z una función real valuada medible definida en $(\mathcal{X}, \mathcal{A})$. Si $\int |Z| \mathrm{d}\alpha < \infty$, entonces $\int |Z| \mathrm{d}P < \infty$ con probabilidad uno y*

$$\mathbb{E}\left[\int Z \mathrm{d}P\right] = \int Z \mathrm{d}\mathbb{E}[P] = \alpha(\mathcal{X})^{-1} \int Z \mathrm{d}\alpha.$$

Demostración.

De la definición dada en (19)

$$\int |Z| \mathrm{d}P = \sum_{j=1}^{\infty} |Z(V_j)| P_j. \tag{20}$$

Entonces utilizando el *Teorema de Convergencia Monótona* y la independencia entre V_j y P_j, al sacar la esperanza de la ecuación anterior tenemos que

$$
\begin{aligned}
\mathbb{E}\left[\int |Z|\mathrm{d}P\right] &= \mathbb{E}\left[\sum_{j=1}^{\infty}|Z(V_j)|P_j\right] \\
&\overset{T.C.M.}{=} \sum_{j=1}^{\infty}\mathbb{E}[|Z(V_j)|P_j] \\
&= \sum_{j=1}^{\infty}\mathbb{E}[|Z(V_j)|]\mathbb{E}[P_j] \\
&= \alpha(\mathcal{X})^{-1}\int |Z|\mathrm{d}\alpha \sum_{j=1}^{\infty}\mathbb{E}[P_j] \\
&= \alpha(\mathcal{X})^{-1}\int |Z|\mathrm{d}\alpha,
\end{aligned}
$$

en donde para la última igualdad hacemos uso del hecho de que para todo $j \geq 1$, $P_j \geq 0$ y $\sum_{j=1}^{\infty} P_j = 1$, con lo que el intercambio entre suma y el operador esperanza es válido. Entonces, $\int |Z|\mathrm{d}P$ es finita con probabilidad uno pues $\int |Z|\mathrm{d}\alpha$ lo es y α es una medida finita. Por lo tanto

$$
\int Z\mathrm{d}P = \sum_{j=1}^{\infty} Z(V_j)P_j,
$$

converge absolutamente de manera casi segura. Para concluir la prueba, sólo falta notar que esta última serie está acotada por (20), la cual es integrable, y entonces utilizando el *Teorema de Convergencia Dominada* y desarrollando de manera análoga a lo hecho en la primera parte de esta prueba, tenemos que

$$
\begin{aligned}
\mathbb{E}\left[\int Z\mathrm{d}P\right] &= \sum_{j=1}^{\infty}\mathbb{E}[Z(V_j)]\mathbb{E}[P_j] \\
&= \alpha(\mathcal{X})^{-1}\int Z\mathrm{d}\alpha.
\end{aligned}
$$

$\diamond$

Este resultado implica en particular que si $(\mathcal{X}, \mathcal{A})$ fuera el espacio medible formado por la recta real y los borelianos de $\mathbb{R}$, y además α tuviera k-ésimo momento finito, entonces con probabilidad uno, P tendría k-ésimo momento finito.

3.2 Aplicación.

Para fines de este documento, hablaremos y desarrollaremos una de las varias aplicaciones existentes en la estadística no paramétrica desde el punto de vista Bayesiano con las herramientas mostradas previamente, nos referimos a la estimación de una función de distribución.

Se recomienda ampliamente la lectura de la sección 5 del artículo de *Ferguson* [Fer73], al lector interesado en conocer a detalle aplicaciones que van desde la estimación de la media de una distribución, hasta el contraste de hipótesis desde el punto de vista de la estadística Bayesiana no paramétrica.

A lo largo de esta sección, tomaremos α como una medida finita σ-aditiva no nula en $(\mathcal{X}, \mathcal{A})$. Escribiremos $P \in \mathscr{D}(\alpha)$ para denotar la frase "P es un *Proceso Dirichlet* en

$(\mathcal{X}, \mathcal{A})$ con parámetro, el argumento entre paréntesis". $\mathscr{B}$ denotará la σ-álgebra de *Borel* en $\mathbb{R}$.

El problema de decisión de la estadística Bayesiana no paramétrica que consideraremos, se describe como sigue. El espacio parametral es el conjunto de todas las medidas de probabilidad P en $(\mathcal{X}, \mathcal{A})$. Se elige una acción a en algún espacio (espacio de acciones posibles) de modo que se incurra en la pérdida, $L(P, a)$ y existe una muestra $X_1, \ldots, X_n$ tomada de P sobre la cual se basa la elección de la acción a. Posteriormente, se busca una regla de *Bayes* con respecto a la distribución apriori, $P \in \mathscr{D}(\alpha)$. Con tal distribución apriori, la distribución posterior de P dadas las observaciones es $\mathscr{D}(\alpha + \sum_{i=1}^{n} \delta_{X_i})$, donde δ_x denota la medida que otorga masa 1 al punto x. Entonces, si podemos encontrar una regla de *Bayes* para el problema con $n = 0$ ("*no-sample problem*"), la regla de *Bayes* para el problema general puede encontrarse reemplazando α con $\alpha + \sum_{i=1}^{n} \delta_{X_i}$. En el siguiente desarrollo primero encontramos la ragla de *Bayes* para $n = 0$ y posteriormente establecemos la regla de *Bayes* para el problema general.

3.2.1 Estimación de una Función de Distribución.

Sea $(\mathcal{X}, \mathcal{A}) = (\mathbb{R}, \mathscr{B})$ y tomemos como espacio de acciones posibles, el espacio de todas las funciones de distribución en $\mathbb{R}$. Definamos a la función de pérdida como sigue

$$L(P, \hat{F}) := \int (F(t) - \hat{F}(t))^2 \mathrm{d}W(t),$$

donde W es una medida finita dada en $(\mathbb{R}, \mathscr{B})$ (función de pesos) y $F(t) = P\big((-\infty, t]\big)$. Si $P \in \mathscr{D}(\alpha)$, entonces $F(t) \sim Beta(\cdot|\alpha\big((-\infty, t]\big), \alpha\big((t, \infty)\big)$ para cada t. La función de riesgo para el problema sin muestra $(n = 0)$ es

$$\mathbb{E}[L(P, \hat{F})] = \int \mathbb{E}[(F(t) - \hat{F}(t))^2] \mathrm{d}W(t),$$

que se minimiza si y sólo si para cada t se minimiza $\mathbb{E}[(F(t) - \hat{F}(t))^2]$, lo cual se logra al tomar $\hat{F}(t) = \mathbb{E}[F(t)]$. Por lo tanto, el problema de *Bayes* para el problema sin muestra es

$$\hat{F}(t) = \mathbb{E}[F(t)] = \frac{\alpha\big((-\infty, t]\big)}{\alpha(\mathbb{R})} =: F_0(t),$$

que representa nuestra suposición apriori de la forma de $F(t)$.

Para una muestra de tamaño n, la regla de *Bayes* implica actualizar con la información observada, lo que nos lleva a

$$\begin{aligned} \hat{F}_n(t|X_1, \ldots, X_n) &= \frac{\alpha\big((-\infty, t]\big) + \sum_{i=1}^{n} \delta_{X_i}\big((-\infty, t]\big)}{\alpha(\mathbb{R}) + n} \\ &= p_n F_0(t) + (1 - p_n) F_n(t|X_1, \ldots, X_n), \end{aligned} \tag{21}$$

con

$$p_n = \frac{\alpha(\mathbb{R})}{\alpha(\mathbb{R}) + n} \qquad \text{y} \qquad F_n(t|X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}\big((-\infty, t]\big),$$

la función de distribución empírica de la muestra.

La regla de *Bayes* (21) es una mezcla de nuestra suposición apriori de F y de la función de distribución empírica con pesos p_n y $(1 - p_n)$ respectivamente. Si $\alpha(\mathbb{R})$ es grande en comparación con n, entonces estamos otorgando poco peso a las observaciones. Si $\alpha(\mathbb{R})$ es pequeño comparado con n, implica que se está dando poco peso a la suposición apriori de F. Uno puede interpretar $\alpha(\mathbb{R})$ como una medida de fé en la suposición apriori de F, medida en unidades de número de observaciones. Conforme $\alpha(\mathbb{R})$ tiende a cero (la *Distribución Dirichlet* apriori "no informativa"), la estimación de *Bayes* converge a la función de distribución empírica.

Cabe resaltar que sin importar cual sea la verdadera función de distribución, la estimación de *Bayes* (21) converge casi seguramente de manera uniforme. Esto puede verificarse aplicando el teorema de *Glivenko-Cantelli* y notando que $p_n \to 0$ conforme $n \to \infty$.

El resultado para estimar una función de distribución k-dimensional, es completamente análogo.

4 Principio de Grandes Desviaciones para Distribuciones Posteriores Dirichlet.

Las *Grandes Desviaciones* son basicamente un conjunto de resultados asintóticos a eventos de baja ocurrencia denominados "eventos raros" así como el conjunto de métodos para derivar dichos resultados. Su surgimiento se atribuye a actuarios escandinavos interesados en el análisis del riesgo en la industria aseguradora [Ess32]. En el apéndice A, se encuentran enunciados algunos resultados importantes de la *Teoría de Grandes Desviaciones* que son utilizados a lo largo del resto de este documento. Para una revisión introductoria al tema, se recomienda ampliamente la lectura del articulo de *Varadhan* [Var08], mientras que para el lector interesado en una revisión más completa se recomienda [DZ93].

El *Principio de Grandes Desviaciones* caracteriza la conducta límite de una familia de medidas de probabilidad sobre un cierto espacio en términos de una *Función Tasa* al encontrar cotas (inferior y superior) exponenciales asintóticas. Los *Procesos Dirichlet* son una familia conjugada de distribuciones apriori en la inferencia estadistica Bayesiana no paramétrica, cuya sucesión de distribuciones posteriores cumple un *Principio de Grandes Desviaciones*. En esta sección demostraremos que si la distribución apriori es *Dirichlet*, entonces la sucesión de distribuciones posteriores satisface un *Principio de Grandes Desviaciones*, y daremos una expresión explícita para su *Función Tasa*. Entonces, si modelamos nuestra incertidumbre mediante *Procesos Dirichlet* apriori, podemos asegurar la conducta límite de sus distribuciones posteriores.

La exposición de esta sección se basa en el trabajo de *A.J. Ganesh y N. O'conell* [GO00], quienes además aplican estos resultados para obtener una fórmula asintótica para la probabilidad de ruina en el problema clásico de *la ruina del jugador*.

Sea $\mathcal{X}$ un espacio de *Hausdorff* con σ-álgebra de *Borel* $\mathscr{B}_{\mathcal{X}}$ y sea μ_n una sucesión de medidas de probabilidad en $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$.

Definición 5. *Una **Función Tasa** es una función en $\mathcal{X}$ que es no negativa y semicontinua por abajo.*

Definición 6. *Se dice que la sucesión μ_n satisface el* Principio de Grandes Desviaciones *(**LDP** por sus siglas en inglés) con función tasa I, si para todo $B \in \mathscr{B}_{\mathcal{X}}$,*

$$-\inf_{x\in B^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{n} \ln\big(\mu_n(B)\big) \leq \limsup_{n\to\infty} \frac{1}{n} \ln\big(\mu_n(B)\big) \leq -\inf_{x\in \bar{B}} I(x).$$

Sea Ω un *espacio polaco* (espacio métrico completo y separable) y denotemos por $\mathcal{M}_1(\Omega)$ al espacio de medidas de probabilidad en Ω. Consideremos una sucesión de variables aleatorias independientes $\{X\}_{k\geq 1}$ que toman valores en Ω con ley común $\mu \in \mathcal{M}_1(\Omega)$ y denotemos con L_n a la medida empírica correspondiente a las primeras n observaciones

$$L_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k},$$

y sea $\mathcal{L}(L_n)$ la ley de L_n. Para $\nu \in \mathcal{M}_1(\Omega)$ se define la distancia de *Kullback-Leibler* o *Entropía Relativa* (relativa a μ) como

$$H(\nu|\mu) = \begin{cases} \int_\Omega \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \ln\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)\mathrm{d}\mu & \nu \ll \mu \\ \infty & c.o.c \end{cases}$$

El postulado del *Teorema de Sanov* dice que la sucesión $\mathcal{L}(L_n)$ satisface un **LDP** en $\mathcal{M}_1(\Omega)$ equipado con la topología τ (topología más fina o fuerte que la débil), con función tasa $H(\cdot|\mu)$ (ver *Dembo y Zeitouni* [DZ93], *Teorema 6.2.10*). *A.J. Ganesh y N. O'conell* [GO99], prueban un inverso a este resultado que aparece naturalmente en el contexto *Bayesiano* para conjuntos finitos, en donde el papel de los argumentos de la distancia de *Kullback-Leibler* o *Entropía Relativa* se invierte en comparación con el *Teorema de Sanov*.

A continuación se prueba un **LDP** para el caso de *Procesos Dirichlet* como distribuciones apriori en un espacio métrico y compacto.

Sabemos que si P es un *Proceso Dirichlet* (dsitribución apriori), entonces dadas las observaciones $X_1, \ldots, X_n$, la distribución posterior es también un *Proceso Dirichlet* pero con parámetro $\alpha + \sum_{i=1}^{n} \delta_{X_i}$, es decir que los *Procesos Dirichlet* son una familia conjugada. Esta propiedad facilita sustancialmente el cálculo de la distribución posterior y es muy útil para el trabajo analítico.

El siguiente teorema establece un **LDP** para la sucesión de distribuciones posteriores $\{\mathscr{D}(\alpha + \sum_{i=1}^{n} \delta_{X_i}), n = 1, 2, \ldots\}$. Tomemos ahora a Ω como un espacio métrico y compacto con σ-álgebra de *Borel* $\mathscr{B}_\Omega$. Sea $\mathcal{M}_1(\Omega)$ el espacio de las medidas de probabilidad en $(\Omega, \mathscr{B}_\Omega)$, y $\mathscr{B}(\mathcal{M}_1(\Omega))$ la σ-álgebra de *Borel* inducida por la topología débil en $\mathcal{M}_1(\Omega)$.

Teorema 5. *Sea α una medida finita no negativa en $(\Omega, \mathscr{B}_\Omega)$ con soporte Ω. Sea μ una medida de probabilidad en $(\Omega, \mathscr{B}_\Omega)$, y $\{X_i\}_{i\geq 1}$ una sucesión Ω-valuada tal que*

$$\frac{1}{n}\sum_{i=1}^{n}\delta_{X_i} \overset{n\to\infty}{\longrightarrow} \mu, \qquad \textit{débilmente.}$$

*Entonces la sucesión de medidas de probabilidad $\mathscr{D}(\alpha + \sum_{i=1}^{n}\delta_{X_i})$, satisface un **LDP** en $\mathcal{M}_1(\Omega)$ equipado con su topología débil y con función tasa $I(\cdot)$ dada por*

$$I(\nu) = H(\mu|\nu),$$

donde $H(\mu|\nu)$ denota la entropía relativa de μ con respecto a ν.

La prueba del resultado anterior se basa en los siguientes tres lemas. La siguiente definición nos ayudará a establecer el primero de ellos.

Sea μ_n un elemento de $\mathcal{M}_1(\Omega)$ con distribución $\mathscr{D}(\alpha + \sum_{i=1}^{n}\delta_{X_i})$. Para funciones acotadas y medibles definimos

$$\Lambda_n(f) := \ln \mathbb{E}\Big[\exp\Big\{\int_\Omega f\mathrm{d}\mu_n\Big\}\Big] \tag{22}$$

Lema 1. *Sea $(A_1, \ldots, A_k)$ una partición medible de Ω y supongamos que el interior de A_i es no-vacío para cada $i = 1, \ldots, k$. Sea f una función acotada y medible con respecto a $\sigma(A_1, \ldots, A_k)$, la σ-álgebra generada por los conjuntos $A_1, \ldots, A_k$. Entonces,*

$$\Lambda(f) := \lim_{n\to\infty}\frac{1}{n}\Lambda_n(nf), \tag{23}$$

existe, es finito y esta dado por

$$\Lambda(f) = \sup_{\nu\in\mathcal{M}_1(\Omega)}\Big\{\int_\Omega f\mathrm{d}\nu - H(\mu|\nu)\Big\}. \tag{24}$$

Demostración.

Tomemos $(A_1, \ldots, A_k)$ como en el enunciado y adicionalmente al supuesto del interior no vacío de cada A_i, supongamos que A_i es un conjunto de μ-continuidad ($i.e.$ $\mu(\partial A_i) = 0$) para cada $i = 1, \ldots, k$.

Sea f como se especifica en el lema, entonces podemos escribir

$$f = \sum_{i=1}^{k} c_i \mathbb{I}_{\{A_i\}}, \tag{25}$$

para algunas constantes c_i.

Según (22) tenemos que

$$\Lambda_n(f) = \ln \mathbb{E}\Big[\exp\Big\{\sum_{i=1}^{k} c_i\mu_n(A_i)\Big\}\Big]. \tag{26}$$

Por lo tanto, como cada A_i tiene interior no vacío y ya que α tiene soporte igual a Ω, podemos definir

$$\alpha_n(A_j) := \alpha(A_j) + \sum_{i=1}^{n} \delta_{X_i}(A_j) > 0, \quad \forall n \in \mathbb{N} \quad y \quad j = 1, \dots, k. \tag{27}$$

De la definición de la *Distribución Dirichlet*, sabemos que

$$(\mu_n(A_1), \dots, \mu_n(A_k)) \sim \left(\frac{Z_n^1}{\sum_{i=1}^k Z_n^i}, \dots, \frac{Z_n^k}{\sum_{i=1}^k Z_n^i} \right),$$

en donde para $i = 1, \dots, k$, las variables aleatorias Z_n^i son independientes y con distribución gama, tales que

$$Z_n^j \sim Gama(\cdot|\alpha_n(A_j), 1),$$

y α_n está definida como en (27). Al evaluar el logaritmo de la función generadora de momentos de Z_n^j, obtenemos

$$\lambda_n^j(\theta) := \ln \mathbb{E}\big[\exp\{\theta Z_n^j\}\big] = \ln\Big[\Big(\frac{1}{1-\theta}\Big)^{\alpha_n(A_j)}\Big] = \begin{cases} -\alpha_n(A_j)\ln(1-\theta) & \text{si } \theta < 1 \\ \infty & c.o.c \end{cases}$$

Como por hipótesis (Teorema 5)

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(A_j) \stackrel{n\to\infty}{\longrightarrow} \mu(A_j), \qquad \text{débilmente},$$

y $\alpha(A_j)/n \stackrel{n\to\infty}{\longrightarrow} 0$ ya que α es una medida finita, implica que al sustituir (27) en $\lambda_n^j(\theta)$ obtenemos

$$\lambda_j(\theta) := \lim_{n\to\infty} \frac{1}{n}\lambda_n^j(\theta) = \begin{cases} -\mu(A_j)\ln(1-\theta) & \text{si } \theta < 1 \\ \infty & c.o.c \end{cases},$$

y por el teorema de *Gärthner-Ellis* (ver apéndice A o *Teorema 2.3.6* [DZ93]), la sucesión de variables aleatorias Z_n^j/n satisface un **LDP** en $\mathbb{R}$ con función tasa λ_j^*, la cual es el dual convexo de λ_j, es decir

$$\lambda_j^*(x) = \sup_{\theta\in\mathbb{R}}\{\theta x - \lambda_j(\theta)\} = \begin{cases} x - \mu(A_j) + \mu(A_j)\ln\big[\frac{\mu(A_j)}{x}\big] & \text{si } x > 0 \\ \infty & c.o.c \end{cases} \tag{28}$$

Como $\{Z_n^j, j = 1, \dots, k\}$ son variables independientes, entonces $\{Z_n^j/n, j = 1, \dots, k\}$ satisface de manera conjunta un **LDP** en $\mathbb{R}^k$ con función tasa $\lambda^*(x) = \sum_{j=1}^k \lambda_j^*(x_j)$, donde $x = (x_1, \dots, x_k)$ y λ_j^* está dada por (28).

Definamos

$$Y_n^j := \frac{Z_n^j}{\sum_{i=1}^k Z_n^i},$$

y ya que $\sum_{i=1}^k Z_n^i$ es estrictamente positiva con probabilidad 1, los mapeos

$$(Z_n^1, \dots, Z_n^k) \mapsto (Y_n^1, \dots, Y_n^k),$$

son continuos casi seguramente para cada n. Se sigue del *Principio de Contracción* (*Teorema 4.2.1* de [DZ93] o Apéndice A) que la sucesión $\{Y_n^j, j = 1, \ldots, k\}$ satisface de manera conjunta un **LDP**, con función tasa I dada por

$$I(y_1, \ldots, y_k) = \text{ínf}\left\{\sum_{j=1}^{k} \lambda_j^*(z_j) \ : \ y_j = \frac{z_j}{\sum_{i=1}^{k} z_i}, \quad j = 1, \ldots, k\right\}. \tag{29}$$

Si $y_j < 0$ para algún j, implica que cualquier z incluido en el ínfimo en (29) debe tener un $z_i < 0$ para algún i y entonces por (28), $I(y) = \infty$. Por otro lado, si $y_j = 0$ para todo j o si $\sum_{i=1}^{n} y_i \neq 1$, entonces no existe $z \in \mathbb{R}^k$ tal que

$$y_j = \frac{z_j}{\sum_{i=1}^{k} z_i},$$

para toda $j = 1, \ldots, k$, y por lo tanto nuevamente $I(y) = \infty$, al ser el ínfimo de un conjunto vacío.

A partir de este punto confinaremos nuestra atención a los $y \in \mathbb{R}^k$ tales que $y \geq 0$ y $\sum_{i=1}^{k} y_i = 1$. Si $z \in \mathbb{R}^k$ es tal que

$$y_j = \frac{z_j}{\sum_{i=1}^{k} z_i}, \qquad \forall \quad j = 1, \ldots, k,$$

entonces podemos escribir $z = \beta y$ para algún $\beta > 0$. De este modo (29) queda como

$$I(y_1, \ldots, y_k) = \text{ínf}_{\beta > 0} \sum_{j=1}^{k} \lambda_j^*(\beta y_j). \tag{30}$$

Al derivar la suma del lado derecho con respecto a β e igualar a cero obtenemos

$$0 = \sum_{j=1}^{k} \left(y_j - \frac{\mu(A_j)}{\beta} \right) = 1 - \frac{1}{\beta},$$

donde para obtener la última igualdad hemos usado que $\sum_{j=1}^{k} y_j = 1$ y que $\sum_{j=1}^{k} \mu(A_j) = 1$ pues μ es una distribución de probabilidad y $A_1, \ldots, A_k$ es una partición medible de Ω.

Como cada λ_j^* es una función convexa (*lema 2.2.5* de [DZ93]), entonces al resolver la última ecuación para β tenemos que el ínfimo de (30) se alcanza en $\beta = 1$, además

$$\begin{aligned} I(y) = \sum_{j=1}^{k} \lambda_j^*(y_j) &= \sum_{j=1}^{k} y_j - \mu(A_j) + \mu(A_j) \ln\left[\frac{\mu(A_j)}{y_j}\right] \\ &= \sum_{j=1}^{k} \mu(A_j) \ln\left[\frac{\mu(A_j)}{y_j}\right]. \end{aligned}$$

La segunda igualdad se sigue de (28) y para la tercera se utiliza el hecho de que tanto y como μ son distribuciones de probabilidad y por lo tanto suman 1.

De lo anterior se sigue que la sucesión de vectores aleatorios $(\mu_n(A_1), \ldots, \mu_n(A_k))$ satisface un **LDP** en $\mathbb{R}^k$ con función tasa

$$I(y) = \begin{cases} \sum_{j=1}^{k} \mu(A_j) \ln\left[\frac{\mu(A_j)}{y_j}\right] & \text{si } y \in \mathbb{R}_+^k \text{ y } \sum_{i=1}^{k} y_i = 1 \\ \infty & c.o.c \end{cases} \tag{31}$$

Observemos que de (25) se cumple que

$$\Big|\int f \mathrm{d}\mu\Big| = \Big|\int \sum_{i=1}^{k} c_i \mathbb{I}_{\{A_i\}}\Big| \le \max_{i \in \{1,\dots,k\}} |c_i|,$$

ya que μ_n es una distribución de probabilidad. Por lo tanto, por el *lema de Varadhan* (Apéndice A o [DZ93] *Teorema 4.3.1*) y del **LDP** para $(\mu_n(A_1), \dots, \mu_n(A_k))$ se tiene que

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n} \Lambda_n(nf) = \sup_{y \in \mathbb{R}^k} \Big\{ \sum_{i=1}^{k} c_i y_i - I(y) \Big\}.$$

Utilizando (31), podemos reescribir $\Lambda(f)$ como

$$\begin{aligned} \Lambda(f) &= \sup_{\nu \in \mathcal{M}_1(\Omega)} \Big\{ \sum_{i=1}^{k} c_i \nu(A_i) - \sum_{i=1}^{k} \mu(A_i) \ln\Big[\frac{\mu(A_i)}{\nu(A_i)}\Big] \Big\} \\ &= \sup_{\nu \in \mathcal{M}_1(\Omega)} \Big\{ \int_\Omega f \mathrm{d}\nu - H_k(\mu|\nu) \Big\}, \end{aligned} \tag{32}$$

donde

$$H_k(\mu|\nu) := \sum_{i=1}^{n_k} \mu(A_i^k) \ln\Big[\frac{\mu(A_i^k)}{\nu(A_i^k)}\Big].$$

A continuación mostraremos que se puede reemplazar $H_k(\mu|\nu)$ por $H(\mu|\nu)$ en (32). Sea $\nu \in \mathcal{M}_1(\Omega)$ arbitraria. Si $\mu(A_i) > 0$ y $\nu(A_i) = 0$ para algún A_i, entonces $\mu \not\ll \nu$ y por lo tanto $H(\mu|\nu)$ y $H_k(\mu|\nu)$ son ambas infinito. Entonces sin pérdida de generalidad podemos suponer que $\mu(A_i) = 0$ siempre que $\nu(A_i) = 0$.

Definamos $\lambda \in \mathcal{M}_1(\Omega)$ somo sigue. Fijemos $\lambda \equiv \nu$ en A_i si $\mu(A_i) = 0$, y si $\mu(A_i) > 0$, tomemos λ absolutamente continua con respecto a μ en A_i, con derivada de *Radon-Nikodym*

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \equiv \frac{\nu(A_i)}{\mu(A_i)} > 0$$

De este modo μ es absolutamente continua con respecto a λ y entonces

$$\begin{aligned} H(\mu|\lambda) &= \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \ln\Big[\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big] \mathrm{d}\lambda = \int_\Omega \mathrm{d}\mu \ln\Big[\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big] \\ &= \sum_{i:\mu(A_i)>0} \int_{A_i} \mathrm{d}\mu \ln\Big[\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big] = \sum_{i:\mu(A_i)>0} \mu(A_i) \ln\Big[\frac{\mu(A_i)}{\nu(A_i)}\Big] \\ &= H_k(\mu|\nu). \end{aligned} \tag{33}$$

Por supuesto,

$$\int_\Omega f \mathrm{d}\nu = \sum_{i=1}^{k} c_i \nu(A_i) = \sum_{i=1}^{k} c_i \lambda(A_i) = \int_\Omega f \mathrm{d}\lambda. \tag{34}$$

Como $\nu \in \mathcal{M}_1(\Omega)$ fue arbitraria, de (33) y (34) podemos concluir que

$$\sup_{\nu\in\mathcal{M}_1(\Omega)} \Big\{ \int_\Omega f \mathrm{d}\nu - H_k(\mu|\nu) \Big\} \leq \sup_{\lambda\in\mathcal{M}_1(\Omega)} \Big\{ \int_\Omega f \mathrm{d}\lambda - H(\mu|\lambda) \Big\},$$

mientras que la desigualdad inversa es consecuencia de (34) y el hecho de que $H_k(\mu|\nu) \leq H(\mu|\nu)$ para toda $\nu \in \mathcal{M}_1(\Omega)$, puesto que el mapeo $x \longmapsto x \ln x$ es convexo en $[0,\infty)$ (o como consecuencia del *Lema 2*, que veremos más adelante). Por lo tanto

$$\sup_{\nu\in\mathcal{M}_1(\Omega)} \Big\{ \int_\Omega f \mathrm{d}\nu - H_k(\mu|\nu) \Big\} = \sup_{\lambda\in\mathcal{M}_1(\Omega)} \Big\{ \int_\Omega f \mathrm{d}\lambda - H(\mu|\lambda) \Big\},$$

con lo que se concluye la prueba del lema afirmando la ecuación (24).

◇

Lema 2. *Sea $A^k = (A_1^k, \ldots, A_{n_k}^k)$ con $k \in \mathbb{N}$ una sucesión de particiones medibles de Ω tales que las correspondientes σ-álgebras, $\sigma(A^k)$ crecen a $\mathscr{B}(\Omega)$, la σ-álgebra de* Borel *en Ω. Entonces para todo $\nu \in \mathcal{M}_1(\Omega)$, tenemos*

$$H(\mu|\nu) = \sup_{k\geq 1} H_k(\mu|\nu) = \lim_{k\to\infty} H_k(\mu|\nu),$$

con

$$H_k(\mu|\nu) := \sum_{i=1}^{n_k} \mu(A_i^k) \ln\Big[\frac{\mu(A_i^k)}{\nu(A_i^k)}\Big].$$

Demostración.

Comenzaremos por enfatizar el hecho de que $H(\mu|\nu)$ es una función creciente de $\sigma(A^k)$ (ver *prop. 15.5* [Gor88]). De modo que se cumple que

$$H(\mu|\nu) \geq \sup_{k\geq 1} H_k(\mu|\nu) = \lim_{k\to\infty} H_k(\mu|\nu),$$

donde $H(\mu|\nu)$ se toma sobre $\mathscr{B}(\Omega)$, y cada $H_k(\mu|\nu)$ es tomada sobre $\sigma(A^k)$ para cada $k \geq 1$. Entonces la desigualdad de arriba se sigue de inmediato ya que $\sigma(A^k) \nearrow \mathscr{B}(\Omega)$, y ésta última es la σ-álgebra más pequeña que contiene a todas las $\sigma(A^k)$.

Denotemos a el $\lim_{k\to\infty} H_k(\mu|\nu)$ por C, una constante que supondremos finita, pues de otro modo tendremos que $\mu \leq \nu$ y entonces ambos lados de la ecuación que queremos probar son infinito y entonces no hay nada que hacer. Por lo tanto, lo que nos resta por demostrar es que

$$H(\mu|\nu) \leq C, \quad C < \infty.$$

Como C es finita, implica que μ es ν-continua en cada $\sigma(A^k)$ y entonces existe una sucesión de funciones $\{f_k\}_{k\geq 1}$ positivas que son $\sigma(A^k)$-medibles tales que $\mu = f_k\nu$ en $\sigma(A^k)$.

A continuación se muestra que $\{f_k\}_{k\geq 1}$ converge en la norma $\mathscr{L}^1(\nu)$ a alguna función $f \geq 0$, $\mathscr{B}(\Omega)$-medible. Primero notemos que $\{f_k\}_{k\geq 1}$ es una sucesión uniformemente ν-integrable pues para $r > 1$

$$0 \leq \nu(f_k \mathbb{I}_{\{f_k \geq r\}}) \leq (\ln r)^{-1} \nu(f_k \ln f_k \mathbb{I}_{\{f_k \geq r\}}),$$

ya que

$$\frac{\ln f_k}{\ln r} \geq 1,$$

si la función indicadora dentro de las integrales de arriba no se anula. Por otro lado

$$(\ln r)^{-1} \nu(f_k \ln f_k \mathbb{I}_{\{f_k \geq r\}}) \leq (\ln r)^{-1} \Big(\sup_{k\geq 1} H_k(\mu|\nu) + 1 \Big) = (\ln r)^{-1}(C+1) < \infty,$$

utilizando que $x \ln x \geq -1$, $x \geq 0$ y la definición de $H_k(\mu|\nu)$. Entonces

$$\nu(f_k \mathbb{I}_{\{f_k \geq r\}}) \leq (\ln r)^{-1}(C+1).$$

Por lo tanto, el lado derecho de ésta desigualdad tiende a cero conforme r tiende a infinito.

De hecho, la sucesión también es de *Cauchy* en $\mathscr{L}^1(\nu)$. Para probar ésto último, supongamos lo contrario. Entonces existe un $\epsilon > 0$ y una sucesión creciente $\{\alpha_n\}_{n\geq 1}$ de números naturales tales que $\nu(|f_{\alpha_{n+1}} - f_{\alpha_n}|) \geq \epsilon$, para toda $n \geq 1$. Sin embargo, ésto es imposible ya que $\{f_{\alpha_n}\}_{n\geq 1}$ es uniformemente integrable y por lo tanto converge en $\mathscr{L}^1(\nu)$, por lo tanto la sucesión es de *Cauchy*.

De este modo, como $\mathscr{L}^1(\nu)$ es un espacio completo concluimos que existe una función $\mathscr{B}(\Omega)$-medible $f \geq 0$, tal que $\lim_{k\to\infty} \nu(|f_k - f|) = 0$. Evidentemente $\mu = f\nu$ en $\mathscr{B}(\Omega)$, con lo que afirmamos que $H(\mu|\nu) \leq C$ ya que al tomar algún entero positivo N y considerar las funciones $g = N \wedge (f \ln f)$ y $g_k = N \wedge (f_k \ln f_k)$. Como f_k converge en probabilidad a f por ser convergente en norma $\mathscr{L}^1(\nu)$. Implica que g_k converge en probabilidad a g conforme $-1 \leq g_k, g \leq N$. Lo que nos dice que $\lim_{k\to\infty} \nu(|g_k - g|) = 0$, y por lo tanto

$$\nu(g) = \lim_{k\to\infty} \nu(g_k) \leq \lim_{k\to\infty} H_k(\mu|\nu) = C,$$

y haciendo tender N a infinito obtenemos que $H(\mu|\nu) \leq C$.

◇

Lema 3. *Para toda función $f : \Omega \mapsto \mathbb{R}$ continua y acotada, el límite en (23) existe y es finito. El mapeo $\Lambda : \mathscr{C}_b(\Omega) \longmapsto \mathbb{R}$ es convexo y continuo. Además*

$$\Lambda(f) = \sup_{\nu \in \mathcal{M}_1(\Omega)} \Big[\int f \mathrm{d}\nu - H(\mu|\nu) \Big].$$

$\mathscr{C}_b(\Omega)$, denota el espacio de todas las funciones continuas y acotadas que van de Ω a $\mathbb{R}$, equipado con la norma supremo $\|f\|_\infty = \sup_{x\in\Omega} |f(x)|$.

Demostración.

Sea $\epsilon > 0$, y $f : \Omega \mapsto \mathbb{R}$ continua y acotada. Entonces dada la definición de f, podemos encontrar $k > 0$ y $g = \sum_{i=1}^{k} c_i \mathbb{I}_{\{A_i\}}$ simple, tal que $\|f - g\|_\infty < \epsilon$. De hecho por la continuidad de f podemos elegir los conjuntos A_i como conjuntos de μ-continuidad con interior no vacío.

Por (22) y el hecho de que cada μ_n es una distribución de probabilidad,

$$\Lambda(nf) = \ln \mathbb{E}\Big[\exp\Big\{\int_\Omega nf \mathrm{d}\mu_n\Big\}\Big] \leq \ln \mathbb{E}\Big[\exp\Big\{\int_\Omega ng \mathrm{d}\mu_n + n\epsilon\Big\}\Big] = n\epsilon + \Lambda_n(ng),$$

de modo que utilizando (23) tenemos que

$$\limsup_{n\to\infty} \frac{1}{n}\Lambda_n(nf) \leq \Lambda(g) + \epsilon.$$

De igual manera

$$\liminf_{n\to\infty} \frac{1}{n}\Lambda_n(nf) \geq \Lambda(g) - \epsilon.$$

Como $\epsilon > 0$ es arbitraria, se sigue que

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n}\Lambda_n(nf)$$

Existe y es finito para todas las funciones $f : \Omega \mapsto \mathbb{R}$ continuas y acotadas. De lo anterior, también podemos concluir que $\Lambda : C_b(\Omega) \longmapsto \mathbb{R}$ es continua, con $|\Lambda(f) - \Lambda(g)| \leq \|f - g\|_\infty$.

Para $f \in \mathscr{L}^\infty(\Omega)$. Definamos

$$H^*(f) = \sup_{\nu \in \mathcal{M}_1(\Omega)} \Big[\int_\Omega f \mathrm{d}\nu - H(\mu|\nu)\Big], \tag{35}$$

es decir que H^* es el conjugado convexo de $H(\mu|\cdot)$.

Como $|\int f \mathrm{d}\nu| \leq \|f\|_\infty$ para toda $\nu \in \mathcal{M}_1(\Omega)$ y $H(\mu|\cdot)$ es no negativa con $H(\mu|\mu) = 0$, entonces $|H^*(f)| \leq \|f\|_\infty$. Ya que H^* es una función convexa con dominio $\mathscr{L}^\infty(\Omega)$, el cual es acotado en la vecindad abierta $\{f : \|f\|_\infty < 1\}$, tenemos que H^* es continua en el interior de su dominio (*Teorema 8.* [Roc74]), el cual es todo $\mathscr{L}^\infty(\Omega)$.

Por último, del *Lema 1* tenemos que H^* y Λ concuerdan en funciones de la forma $f = \sum_{i=1}^{n} c_i \mathbb{I}_{A_i}$, donde la partición $\{A_i\}_{i=1}^{k}$ de Ω tiene a cada A_i como conjunto de μ-continuidad con interior no vacío. Tales funciones son densas en $\mathscr{C}_b(\Omega)$. Previamente hemos demostrado que Λ es continua en $\mathscr{C}_b(\Omega)$ y que H^* es continua en $\mathscr{L}^\infty(\Omega) \supseteq \mathscr{C}_b(\Omega)$ de donde se sigue que $\Lambda = H^*$ en $\mathscr{C}_b(\Omega)$ y consecuentemente se tiene que Λ es convexa.

$\diamond$

Una vez demostrados los tres lemas previamente expuestos, nos encontramos en condiciones de poder probar el *Teorema* principal de ésta sección.

Demostración. **(Teorema 5 sección 4)**

Del *Lema 3* tenemos que Λ es el conjugado convexo de $H(\mu|\cdot)$, pero $H(\mu|\cdot)$ es a su vez convexa y semi-continua por abajo en la topología débil (*Lema 1.4.3* de [DR97]), además sabemos que $\mathcal{M}_1(\Omega)$ es un espacio polaco ya que Ω lo es por hipótesis. Entonces, $H(\mu|\cdot)$ y $\Lambda(\cdot)$ son duales convexos uno del otro.

La cota superior de las grandes desviaciones para subconjuntos compactos de $\mathcal{M}_1(\Omega)$ se sigue del *Teorema 4.5.3* de [DZ93], pero Ω es un espacio compacto por hipótesis, entonces $\mathcal{M}_1(\Omega)$ también es compacto en la topología débil, por lo tanto la cota superior de las grandes desviaciones se tiene para todo conjunto cerrado en $\mathcal{M}_1(\Omega)$.

A continuación probaremos la cota inferior de las grandes desviaciones. La topología débil en $\mathcal{M}_1(\Omega)$ está generada por los conjuntos

$$U_{\phi,x,\delta} = \Big\{\nu \in \mathcal{M}_1(\Omega) \;:\; \Big|\int_\Omega \phi \mathrm{d}\nu - x\Big| < \delta\Big\}, \quad \phi \in \mathscr{C}_b(\Omega),\ x \in \mathbb{R},\ \delta > 0.$$

Dado este conjunto y $\epsilon > 0$, podemos encontrar una sucesión de particiones medibles $A^k = (A_1^k, \dots, A_{n_k}^k)$ de Ω, y una sucesión de funciones simples ϕ_k que sean medibles con respecto a $\sigma(A^k)$, con las siguientes propiedades:

- Las σ-álgebras $\sigma(A^k)$ crecen a $\mathscr{B}(\Omega)$, la σ-álgebra de Borel en Ω.
- Para toda k y toda $i \in \{1, \dots, n_k\}$ es un conjunto de μ-continuidad con interior no vacío.
- Para algún $K > 0$ y todo $k > K$, $\|\phi_k - \phi\|_\infty < \epsilon$.

Por cuestiones estéticas de la prueba, supondremos que $\epsilon < \delta/3$. De modo que

$$\mathbb{P}\{\mu_n \in U_{\phi,x,\delta}\} \geq \mathbb{P}\Big\{\Big|\int_\Omega \phi_k \mathrm{d}\mu_n - x\Big| < \delta - \epsilon\Big\}, \qquad \forall\ k < K, \tag{36}$$

dada la monotonía de la medida de probabilidad $\mathbb{P}$.

Sea $\phi_k = \sum_{i=1}^{n_k} c_i^k \mathbb{I}_{A_i^k}$, entonces $\int_\Omega \phi_k \mathrm{d}\mu_n = \sum_{i=1}^{n_k} c_i^k \mu_n(A_i^k)$. En la prueba del *Lema 1* (ver Ecuación (31)), se muestra que la sucesión $(\mu_n(A_1^k), \dots, \mu_n(A_{n_k}^k))_{n\geq 0}$ satisface un **LDP** con función I_k dada por

$$I_k(y_1, \dots, y_{n_k}) = \begin{cases} \sum_{j=1}^{n_k} \mu(A_j) \ln\left[\frac{\mu(A_j)}{y_j}\right] & \text{si } y \in \mathbb{R}_+^{n_k} \text{ y } \sum_{i=1}^{n_k} y_i = 1 \\ +\infty & \text{c.o.c} \end{cases}$$

Del principio de contracción (*Teorema 4.2.1* de [DZ93] o Apéndice A) se sigue que $\sum_{i=1}^{n_k} c_i^k \mu_n(A_i^k)$ satisface un **LDP** con función tasa J_k dada por

$$\begin{aligned} J_k(x) &= \text{ínf}\{I_k(y) \;:\; \sum_{i=1}^{n_k} c_i^k y_i = x\} \\ &= \text{ínf}\Big\{H_k(\mu|\nu) \;:\; \nu \in \mathcal{M}_1(\Omega), \int_\Omega \phi_k \mathrm{d}\nu = x\Big\}. \end{aligned}$$

En particular, se garantiza la obtención de la siguiente cota inferior

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\Big\{\Big|\int_\Omega \phi_k \mathrm{d}\mu_n - x\Big| < \delta-\epsilon\Big\} &\geq -\inf\{J_k(y)\ :\ |y-x|<\delta-\epsilon\}\\ &= -\inf\Big\{H_k(\mu|\nu)\ :\ \nu\in\mathcal{M}_1(\Omega), \Big|\int_\Omega \phi_k\mathrm{d}\nu - x\Big| < \delta-\epsilon\Big\}.\end{aligned} \tag{37}$$

Notando que $\|\phi-\phi_k\|_\infty < \epsilon$ para todo $k > K$, entonces para toda $\nu\in\mathcal{M}_1(\Omega)$ se tiene que

$$\Big|\int_\Omega \phi\mathrm{d}\nu - x\Big| < \delta - 2\epsilon \quad \Rightarrow \quad \Big|\int_\Omega \phi_k\mathrm{d}\nu - x\Big| < \delta-\epsilon \quad \forall\ k > K.$$

Tomando esto último en cuanta y utilizando las ecuaciones (36) y (37), afirmamos que

$$\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\{\mu_n\in U_{\phi,x,\delta}\} \geq -\inf\Big\{H_k(\mu|\nu)\ :\ \Big|\int_\Omega \phi\mathrm{d}\nu - x\Big| < \delta-2\epsilon\Big\}.$$

Por lo tanto, utilizando el *Lema 2* se tiene

$$\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\{\mu_n\in U_{\phi,x,\delta}\} \geq -\inf\Big\{H(\mu|\nu)\ :\ \Big|\int_\Omega \phi\mathrm{d}\nu - x\Big| < \delta-2\epsilon\Big\},$$

y ya que $\epsilon > 0$ es arbitraria, entonces al hacer tender ϵ a cero obtenemos que

$$\liminf_{n\to\infty}\frac{1}{n}\ln\mathbb{P}\{\mu_n\in U_{\phi,x,\delta}\} \geq -\inf\Big\{H(\mu|\nu)\ :\ \Big|\int_\Omega \phi\mathrm{d}\nu - x\Big| < \delta\Big\},$$

que es la cota inferior de las grandes desviaciones deseada para el conjunto $U_{\phi,x,\delta}$, con función tasa $H(\mu|\cdot)$.

Por lo tanto se ha establecido la cota inferior de las grandes desviaciones para una base de la topología débil en $\mathcal{M}_1(\Omega)$, y en consecuencia para todos los conjuntos abiertos en ésta topología. Al combinar éste último resultado con la cota superior encontrada con anterioridad, se concluye el resultado de *Teorema*.

$\diamond$

Notemos que no hay pérdida de generalidad en el supuesto de que el soporte de α sea Ω. De hecho, si el soporte de la medida apriori α fuera algún subconjunto Ω_1 más pequeño, entonces ya que la medida posterior asigna masa cero fuera de Ω_1, implica que nos podemos confinar al conjunto cerrado Ω_1, el cual es también un espacio métrico y compacto.

5 Mezclas de Distribuciones Apriori Conjugadas y Grandes Desviaciones para Probabilidades de Cruce de Nivel.

La estiamación de probabilidades de cruce de nivel se ha convertido en un tema de gran relevancia en muchos campos de la *Teoría de Riesgo* conectada con problemas de seguros, teoría de colas entre muchos otros. Esta sección está enfocada en mostrar estimaciones de la probabilidad de que un proceso estocástico real-valuado cruce un cierto nivel positivo en un contexto *Bayesiano* basado en el principio de grandes desviaciones.

Para el análisis *Bayesiano* elegimos una mezcla finita de distribuciones apriori conjugadas para modelar la incertidumbre que se tiene sobre los parámetros desconocidos de un *Proceso Poisson Compuesto* con saltos hacia arriba y deriba negativa. Las estimaciones de probabilidades de cruce de nivel tratadas en este documento son derivadas como consecuencia del principio de grandes desviaciones para distribuciones posteriores. Esta es la idea de *C. Macci y L. Petrella* [MP06], quienes además presentan el análisis para el *Movimiento Browniano* con deriva real y precisión positiva. Posteriormente, en 2010 *C. Macci* [Mac10] extiende un poco estos resultados con aplicaciones a la *Teoría de Riesgo*.

Comencemos por conocer el problema de manera formal. Supongamos que Z_t^θ es un proceso estocástico real-valuado que comienza en cero y tiene un parámetro desconocido $\theta \in \Theta$ y definamos a $\tau_Q(\theta)$ como el primer tiempo en el cual el proceso Z_t^θ alcanza un cierto nivel positivo Q, es decir

$$\tau_Q(\theta) = \text{ínf}\{t \geq 0 \ : \ Z_t^\theta \geq Q\}.$$

Como el verdadero valor de θ es desconocido, uno puede desear fijar la probabilidad de cruzar un cierto nivel positivo basados en la experiencia histórica. Esto nos lleva a concíderar lo que comunmente se denomina como *Probabilidad de Cruce de Nivel Predictiva*.

$$\int_\Theta p(Q,\theta)\pi(\mathrm{d}\theta \mid [datos]_n), \tag{38}$$

donde $p(Q,\theta) = \mathbb{P}\{\tau_Q(\theta) < \infty\}$ es la probabilidad de que el proceso con parámetro $\theta \in \Theta$ cruce el nivel positivo Q y $\pi(\cdot \mid [datos]_n)$ representa la distribución posterior de θ dada una muestra aleatoria de Z_t^θ.

Para estimar (38) se elige una mezcla finita de distribuciones apriori conjugadas para modelar la incertidumbre sobre el parámetro desconocido θ de nuestro proceso estocástico Z_t^θ. Este enfoque nos puede llevar a un análisis estadístico completo capturando toda la incertidumbre relacionada con la estimación de la probabilidad de cruce de nivel.[3].

[3]Otras aproximaciones basadas en estimaciones de $p(Q,\hat{\theta})$, dode $\hat{\theta}$ es una estimación de θ, pueden llevar a inferencias muy erradas [GGOP98]

En la *Teoría de Riesgo*, el proceso Poisson Compuesto con saltos hacia arriba y deriva negativa, es el proceso que describe el exceso de reclamaciones para el modelo de *Cramér-Lundberg* (ver [EKM97] o [Asm00]). Este último tiene sus orígenes en la tesis doctoral de *Filip Lundberg* con la que se funda la *Teoría de Riesgo Actuarial*. Posteriormente *Harald Cramér* retoma las ideas originales de *Lundberg* y las pone en el contexto de los procesos estocásticos estableciendo lo que hasta nuestros días se conoce formalmente como *Teoría de Riesgo*. Para mayores detalles sobre esta teoría se recomiendan los textos de *Asmussen* [Asm00] y *Embrechts et al.* [EKM97].

El objetivo de esta sección es presentar un análisis Bayesiano asintótico para probabilidades de cruce nivel basado en las grandes desviaciones. En particular se muestra un principio de grandes desviaciones para las distribuciones posteriores conforme n tiende a infinito bajo el supuesto de convergencia de alguna estadística suficiente adecuada hacia algún valor límite.

Haremos referencia al teorema principal de la sección anterior (recopilado de [GO00]), cuando tratemos con el proceso Poisson Compuesto y elijamos un proceso Dirichlet apriori [Fer73] para la ley común (desconocida) de los saltos. Para la parte relacionada con la *Teoría de Riesgo* haremos referencia a los textos de *Asmussen* [Asm00] y *Embrechts et al.* [EKM97].

Partiendo de un principio de grandes desviaciones para las distribuciones posteriores, se derivarán dos tipos de estimaciones de las probabilidades de cruce de nivel predictivas utilizando el *Lema de Varadhan* (apéndice A o [DZ93]). Una de ellas, para cuando el nivel nivel Q se va a infinito fijando $p(Q,\theta) = p(qn,\theta)$ para algún $q > 0$ fija. La otra, se inspira en lo que se conoce como "*Slow Markov Limit*" (*Asmussen y Nielsen Ejemplo 1 y Teorema 2* [AN95]) fijando $p(Q,\theta) = p(q,\theta(n))$ para alguna $q > 0$ fija y $\theta(n)$ elegida de manera adecuada. El "*Slow Markov Limit*" para el proceso Poisson Compuesto, consiste en que la intensidad del proceso Poisson subyacente diverja a la misma tasa a la que los saltos se van a cero.

Retomando la notación utilizada en la sección anterior, diremos que una función tasa I es "buena", si todos los conjuntos de nivel

$$\{\omega \in \Omega \ : \ I(\omega) \leq \vartheta\}, \qquad \vartheta > 0$$

son conjuntos compactos.

Utilizando la definición de entropía relativa o distancia de *Kullback-Leibler* expuesta con anterioridad, es inmediato calcular la entropía relativa de dos distribuciones gama, $G_{\alpha,\beta_1} := Gama(\cdot|\alpha,\beta_1)$ y $G_{\alpha,\beta_2} := Gama(\cdot|\alpha,\beta_2)$,

$$\begin{aligned} H(G_{\alpha,\beta_1}|G_{\alpha,\beta_2}) &= \int_0^\infty \ln\left(\frac{\frac{\beta_1^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_1\theta}}{\frac{\beta_2^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_2\theta}}\right)\frac{\beta_1^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta_1\theta}\mathrm{d}\theta \\ &= \alpha\left(\frac{\beta_2}{\beta_1} - 1 - \ln\left[\frac{\beta_2}{\beta_1}\right]\right), \end{aligned}$$

y la entropía relativa de dos distribuciones Poisson, $P_{\lambda_1} := Poisson(\cdot|\lambda_1)$ y $P_{\lambda_2} := Poisson(\cdot|\lambda_2)$,

$$
\begin{aligned}
H(P_{\lambda_1}|P_{\lambda_2}) &= \sum_{k\geq 0} \ln\left(\frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}}{\frac{\lambda_2^k}{k!}e^{-\lambda_2}}\right)\frac{\lambda_1^k}{k!}e^{-\lambda_1} \\
&= \lambda_1 \ln\left[\frac{\lambda_1}{\lambda_2}\right] - \lambda_1 + \lambda_2
\end{aligned}
$$

Las siguientes dos ecuaciones serán de gran utilidad para el desarrollo de esta sección[4].La primera de ellas relaciona la entropía relativa de dos distribuciones gama con la de dos distribuciones Poisson.

$$
H(P_{\hat{\lambda}}|P_{\lambda}) = \hat{\lambda} H(G_{1,\hat{\lambda}}|G_{1,\lambda}), \qquad \hat{\lambda}, \lambda > 0. \tag{39}
$$

Mientras que la segunda, servirá cuando se aplique el *Teorema de Gärtner Ellis* a las distribuciones gama.

$$
\sup_{\gamma<\beta}\left\{\gamma\theta - \alpha\ln\left[\frac{\beta}{\beta-\gamma}\right]\right\} = \left\{\begin{array}{ll} \alpha\Big(\beta\frac{\theta}{\alpha} - 1 - \ln\left[\beta\frac{\theta}{\alpha}\right]\Big) = H(G_{\alpha.\frac{\alpha}{\beta}}|G_{\alpha,\theta}) & \text{Si } \theta > 0 \\ \infty & \text{Si } \theta \leq 0 \end{array}\right. , \tag{40}
$$

donde $\alpha, \beta > 0$ y $\gamma = \beta - \frac{\alpha}{\theta}$ maximiza el lado izquierdo de la ecuación.

5.1 Elección de la Distribución Apriori para los Parámetros Desconocidos.

La elección de una mezcla de distribuciones apriori conjugadas para modelar la incertidumbre sobre el parámetro θ del proceso $\{Z_t^\theta\}$, se basa en que éstas son lo suficientemente flexibles para representar una amplia variedad de supuestos apriori sobre nuestro parámetro en cuestión. De hecho *Diaconis y Ylvisaker* [DY85] y *Dalab y Hall* [DH83] prueban que dicha mezcla puede aproximar cualquier distribución de manera arbitraria en función de la exactitud deseada. No obstante, dicha exactitud puede requerir de muchas distribuciones en la mezcla. Por otro lado, un pequeño número de ellas puede capturar una amplia variedad de formas y características de los supuestos apriori.

Sea θ el parámetro desconocido del proceso Z_t^θ. Consideremos a π como la distribución apriori de θ, la cual puede escribirse como una mezcla finita de distribuciones apriori conjugadas π_i,

$$
\pi = \sum_{i=1}^{k} p_i \pi_i, \quad k \text{ un entero positivo}
$$

donde $p_1, \ldots, p_k \geq 0$ y tales que $\sum_{i=1}^k p_i = 1$ son pesos conocidos.

[4]Su importancia radíca en la sencillez del algebra que debe realizarse para obtener las relaciones mostradas.

La implementación del teorema de *Gärtner Ellis* para derivar el principio de grandes desviaciones para las distribuciones posteriores, se preserva bajo mezclas finitas de distribuciones apriori. Este resultado es consecuencia del siguiente lema, el cual se prueba como un resultado más general que involucra cualquier muestra finita de distribuciones apriori que pueden ser no conjugadas.

Lema 4. *Sean $k \geq 1$, $p_1, \ldots, p_k \geq 0$, tales que $\sum_{i=1}^{k} p_i = 1$, cualesquiera distribuciones apriori $\pi_1, \ldots, \pi_k$ y su mezcla $\pi = \sum_{i=1}^{k} p_i \pi_i$. Denotemos a las distribuciones posteriores correspondientes a $\pi_1, \ldots, \pi_k, \pi$ como $\pi_1(\cdot|[datos]_n), \ldots, \pi_k(\cdot|[datos]_n), \pi(\cdot|[datos]_n)$ respectivamente. Aún más, sea f una función medible y supongamos que existe el límite*

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n} \ln \int e^{nf(\theta)} \pi_i(\mathrm{d}\theta|[datos]_n),$$

con $\Lambda(f) \in (-\infty, \infty]$ la cual no depende de $i \in \{1, \ldots, k\}$. Entonces

$$\Lambda(f) := \lim_{n\to\infty} \frac{1}{n} \ln \int e^{nf(\theta)} \pi(\mathrm{d}\theta|[datos]_n), \tag{41}$$

Demostración.

Utilizando el hecho de que la distribución podterior $\pi(\cdot|[datos]_n)$ puede exprezarse como mezcla finita adecuada de las distribuciones $\pi_1(\cdot|[datos]_n), \ldots, \pi_k(\cdot|[datos]_n)$ (ver *O'Hagan y Foster* [OF04]) con pesos que dependen de $p_1, \ldots, p_k$ y los datos $[datos]_n$. De manera más precisa

$$\pi(\cdot|[datos]_n) = \sum_{i=1}^{k} p_i([datos]_n)\pi_i(\cdot|[datos]_n),$$

donde $p_i([datos]_n) \geq 0$ para $i = 1, \ldots, k$, y tales que $\sum_{i=1}^{k} p_i([datos]_n) = 1$. Entonces

$$\int e^{nf(\theta)} \pi(\mathrm{d}\theta|[datos]_n) = \sum_{i=1}^{k} p_i([datos]_n) \int e^{nf(\theta)} \pi(\mathrm{d}\theta|[datos]_n).$$

Consideremos los dos posibles casos por separado. El primero de ellos cuando $\Lambda(f) < \infty$. Para toda $\epsilon > 0$ existe un $N \in \mathbb{N}$ tal que para toda $n \geq N$ se tiene que

$$e^{n(\Lambda(f)-\epsilon)} < \int e^{nf(\theta)} \pi_i(\mathrm{d}\theta|[datos]_n) < e^{n(\Lambda(f)+\epsilon)} \qquad i \in \{1, \ldots, k\}.$$

Entonces, al multiplicar por $p_i([datos]_n)$ y tomar la suma sobre $i \in \{1, \ldots, k\}$ obtenemos

$$e^{n(\Lambda(f)-\epsilon)} < \int e^{nf(\theta)} \pi(\mathrm{d}\theta|[datos]_n) < e^{n(\Lambda(f)+\epsilon)}.$$

Por lo tanto la ecuación (41) se cumple.

Si $\Lambda(f) = \infty$, entonces para todo $M > 0$ existe $N \in \mathbb{N}$, tal que para todo $n \geq N$ se cumple que

$$\int e^{nf(\theta)} \pi_i(\mathrm{d}\theta|[datos]_n) > e^{nM}, \qquad \forall\ i \in \{1, \ldots, k\}$$

de modo que al multiplicar por $p_i([datos]_n)$ y sumar sobre $i \in \{1, \dots, k\}$ obtenemos

$$\int e^{nf(\theta)}\pi(\mathrm{d}\theta|[datos]_n) > e^{nM}, \qquad \forall\ i \in \{1, \dots, k\}$$

y por lo tanto (41) se cumple.

$\diamond$

5.2 Análisis Bayesiano para el Proceso Poisson Compuesto en el contexto de Teoría de Riesgo.

Sea Z_t^θ un proceso Poisson Compuesto con saltos hacia arriba y deriba negativa, de manera más precisa escribamos $Z_t^\theta = \sum_{k=1}^{N_t} B_k - ct$, donde N_t es un proceso Poisson[5] Compuesto con intensidad λ. $\{B_k\}_{k\geq 1}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, positivas e independientes del proceso N_t; $c > 0$ es una constante. Omitimos el caso trivial en que $c \leq 0$, ya que Z_t^θ cruza cualquier nivel positivo con probabilidad 1.

La ley y función generadora de momentos en común de las variables aleatorias B_k, $k \geq 1$. Serán denotadas por ℓ y M_ℓ respectivamente. Aún más, representaremos N_t como $N_t = \sum_{k\geq 1} \mathbb{I}_{\{T_1+\dots+T_k \leq t\}}$, donde $\{T_k\}_{k\geq 1}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución exponencial con tasa λ.

El parámetro θ desconocido es $\theta = (\lambda, \ell)$ y abusando un poco de la notación, denotemos como $\frac{1}{n}\ell$ a ala ley de las variables aleatorias $\{\frac{B_k}{n}\}_{k\geq 1}$ y elijamos $\theta(n) = (n\lambda, \frac{1}{n}\ell)$ para el "*Slow Markov Walk Limit*" [AN95].

Antes de mostrar el **LDP** para las distribuciones posteriores y evaluar la correspondiente probabilidad de cruce de nivel observamos lo siguiente.

Si definimos $w(\lambda, \ell)$ como

$$w(\lambda, \ell) = \sup\{\gamma \geq 0\ :\ \lambda(M_\ell(\gamma) - 1) - c\gamma \leq 0\},$$

que no es otra cosa más, que el máximo sobre los valores posibles de $\gamma \geq 0$ (coeficiente de ajuste) de la *Ecuación de Lundberg* ($\lambda(M_\ell(\gamma) - 1) - c\gamma \leq 0$, ver [Asm00] o [EKM97] para más detalle) .

Una identidad útil y de inmediata verificación dada la definición de $w(\lambda, \ell)$ es:

$$w\Big(n\lambda, \frac{1}{n}\ell\Big) = nw(\lambda, \ell) \tag{42}$$

En las siguientes secciones nos enfocaremos en hacer una estimación de la probabilidad de cruce de nivel cuando la ley ℓ está concentrada en $[0, M]$ para alguna $M > 0$ finita. Haremos lo mismo para cuando $\ell \overset{D}{=} \exp(\cdot|\beta)$, es decir que ℓ es la distribución exponencial con tasa β. En todos los casos a estudiar, afirmaremos que se cumple la *Desigualdad de Lundberg*.

$$\mathbb{P}\{\tau_Q(\lambda, \ell) < \infty\} \leq e^{-w(\lambda,\ell)Q} \tag{43}$$

[5]Proceso Poisson Homogeneo.

(ver *Asmussen*, [Asm00].)

Cuando ℓ está concentrada en $[0, M]$, podemos considerar un refinamiento de la *Desigualdad de Lundberg* de la siguiente manera

$$e^{-w(\lambda,\ell)(Q+M)} \leq \mathbb{P}\{\tau_Q(\lambda,\ell) < \infty\} \leq e^{-w(\lambda,\ell)Q} \tag{44}$$

Utilizando esta última ecuación y (43), tenemos que

$$e^{-nw(\lambda,\ell)(q+\frac{M}{n})} \leq \mathbb{P}\{\tau_{nq}(\lambda,\ell) < \infty\}, \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\} \leq e^{-nw(\lambda,\ell)q},$$

la cual es trivial para $\mathbb{P}\{\tau_{nq}(\lambda,\ell) < \infty\}$, mientras que para $\mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}$ tomamos en cuenta (43) y (44) con $\frac{1}{n}\ell$ y $\frac{M}{n}$ en lugar de ℓ y M respectivamente.

Por lo tanto, para toda $\epsilon > 0$, existe $n_\epsilon \geq 1$ tal que para toda $n \geq n_\epsilon$

$$e^{-nw(\lambda,\ell)(q+\epsilon)} \leq \mathbb{P}\{\tau_{nq}(\lambda,\ell) < \infty\}, \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\} \leq e^{-nw(\lambda,\ell)q}, \tag{45}$$

Es importante remarcar que si ℓ no está concentrada en un conjunto acotado, las cotas inferiores de (44) y (44) no se tienen.

Para el caso $\ell \overset{D}{=} \exp(\cdot|\beta)$ es fácil probar que

$$w(\lambda, \exp\{\beta\}) = \frac{\text{máx}\{\beta c - \lambda, 0\}}{c} = \frac{-\text{mín}\{\lambda - \beta c, 0\}}{c},$$

y entonces se tiene la fórmula cerrada

$$\mathbb{P}\{\tau_Q(\lambda, \exp\{\beta\}) < \infty\} = \left(1 - \frac{w(\lambda, \exp\{\beta\})}{\beta}\right)e^{-w(\lambda,\exp\{\beta\})Q} \tag{46}$$

(*Embrechts et al.*, [EKM97]).

En la siguiente sección se prueba el **LDP** para las distribuciones posteriores de θ y se evalua la probabilidad de cruce de nivel predictiva para los siguientes casos:

- La distribución ℓ es desconocida y concentrada en $[0, M]$, para algún $M > 0$.
- La distribución ℓ es conocida y concentrada en $[0, M]$, para algún $M > 0$.
- La distribución $\ell \overset{D}{=} \exp(\cdot|\beta)$ con $\beta = k\lambda$, donde λ es desconocida y $k > 0$ conocida.

5.3 Distribución ℓ Desconocida y Concentrada en $[0, M]$, con $M > 0$.

Sea $Z_t^{(\lambda,\ell)}$ el proceso Poisson Compuesto descrito anteriormente y supongamos que ℓ es la distribución común de las variables aleatorias $\{B_k\}_{k\geq 1}$ y que se concentra en $[0, M]$. Denotaremos a la familia de todas las medidas de probabilidad en $[0, M]$ por $\mathcal{M}_1([0, M])$, donde supondremos que $\mathcal{M}_1([0, M])$ está equipada con la topología de convergencia débil y denotaremos por $\mathcal{M}_+([0, M])$ a la familia de medidas positivas en $[0, M]$.

Elijamos a $\pi \otimes \mathscr{D}(\mu)$ como la distribución apriori para $\theta = (\lambda, \ell)$, donde

$$\pi = \sum_{i=1}^{k} p_i G_{\alpha(i),\beta(i)},$$

es una mezcla finita de distribuciones gama conjugadas apriori para λ y $\mathscr{D}(\mu)$ el proceso *Dirichlet* apriori para ℓ con parámetro $\mu \in \mathcal{M}_+([0, M])$.

El soporte de $\mathscr{D}(\mu)$ se denotará como $supp(\mathscr{D}(\mu))$. Para el análisis Bayesiano del principio de grandes desviaciones tomaremos una muestra de tamaño n

$$[datos]_n = \big((B_1, T_1), \ldots, (B_n, T_n)\big),$$

y la estadística suficiente formada por la media empírica de $\{T_1, \ldots, T_n\}$ y la ley empírica de $\{B_1, \ldots, B_n\}$,

$$(\bar{\ell}_n, \bar{T}_n) := \left(\frac{1}{n}\sum_{k=1}^{n} \delta_{B_k}, \frac{1}{n}\sum_{k=1}^{n} T_k\right).$$

El sigueinte lema proporciona el **LDP** para las distribuciones posteriores sobre el parámetro θ.

Lema 5. *Sean $k \geq 1$ y $p_1, \ldots, p_k \geq 0$ tales que $\sum_{i=1}^{k} p_i$. Consideremos a $\pi \otimes \mathscr{D}(\mu)$ como la distribución apriori sobre (λ, ℓ), donde $\pi = \sum_{i=1}^{k} p_i G_{\alpha(i),\beta(i)}$. Supongamos también que $\bar{T}_n \to \frac{1}{\hat{\lambda}}$ y $\bar{\ell}_n \to \hat{\ell} \in supp(\mathscr{D}(\mu))$ conforme $n \to \infty$. Entonces $\pi \otimes \mathscr{D}_\mu(\cdot|[datos]_n)$[6] satisface el **LDP** con función tasa buena*

$$I_{\hat{\lambda},\hat{\ell}}(\lambda, \ell) := \begin{cases} H(G_{1,\hat{\lambda}}|G_{1,\lambda}) + H(\hat{\ell}|\ell) & si\ \lambda > 0\ y\ \ell \in supp(\mathscr{D}_\mu) \\ \infty & c.o.c \end{cases}$$

Demostración.

Como las sucesiones $\{B_k\}_{k\geq 1}$ y $\{T_k\}_{k\geq 1}$ son independientes una de la otra, tenemos que $\pi \otimes \mathscr{D}_\mu(\cdot|[datos]_n)$ es la medida producto entre $\pi(\cdot|T_1, \ldots, T_n)$ y $\mathscr{D}_\mu(\cdot|[datos]_n)$. Por otro lado, sabemos que $\mathscr{D}_\mu(\cdot|B_1, \ldots, B_n) =: \mathscr{D}_{\mu_n}$, donde $\mu_n = \mu + n\hat{\ell}_n$ (ver sección 3, Teorema 1. o bien [Fer73]).

Aún más, sabemos que el proceso $\mathscr{D}_{\mu_n}$ satisface un **LDP** con función tasa buena

$$J^{(2)}_{\hat{\ell}}(\ell) = \begin{cases} H(\hat{\ell}|\ell) & \text{si } \ell \in supp(\mathscr{D}_\mu) \\ \infty & c.o.c \end{cases}$$

(sección 4, Teorema 5. o bien [GO00]).

De tal modo que para completar la prueba, debemos demostrar que $\pi(\cdot|T_1, \ldots, T_n)$ satisface un **LDP** con función tasa buena dada por

$$J^{(1)}_{\hat{\lambda}}(\lambda) = \begin{cases} H(G_{1,\hat{\lambda}}|G_{1,\lambda}) & \text{si } \lambda > 0 \\ \infty & \text{si } \lambda \leq 0 \end{cases}$$

[6] Donde redefinimos a $\mathscr{D}(\mu)$ como $\mathscr{D}_\mu$ por conveniencia para nuestra notación en lo subsecuente.

De hecho, si este es el caso, podremos concluir que $\pi \otimes \mathscr{D}_\mu(\cdot|[datos]_n) = \pi(\cdot|T_1, \ldots, T_n) \otimes \mathscr{D}_\mu(\cdot|B_1, \ldots, B_n)$ satisface el **LDP** con función tasa buena

$$I_{\hat{\lambda},\hat{\ell}}(\lambda, \ell) := J^{(1)}_{\hat{\lambda}}(\lambda) + J^{(2)}_{\hat{\ell}}(\ell).$$

Para probar un **LDP** de $\pi(\cdot|T_1, \ldots, T_n)$, recordemos que al elegir a $G_{\alpha,\beta}$ como la conjugada apriori para λ, entonces $G_{\alpha,\beta}(\cdot|T_1, \ldots, T_n) =: G_{\alpha_n,\beta_n}$ donde $\alpha_n = \alpha + n$, $\beta_n = \beta + n\bar{T}_n$ (ver [DeG70]). Por lo tanto

$$\begin{aligned}
\Lambda_{\hat{\lambda}}(\gamma) &= \lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty e^{n\gamma\lambda} G_{\alpha,\beta}(\mathrm{d}\lambda|T_1, \ldots, T_n) \\
&= \lim_{n\to\infty} \begin{cases} \frac{\alpha_n}{n} \ln\left[\frac{\beta_n}{\beta_n - n\gamma}\right] & \text{si } n\gamma < \beta_n \\ \infty & \text{si } n\gamma \geq \beta_n \end{cases} \\
&= \begin{cases} \ln\left[\frac{\frac{1}{\hat{\lambda}}}{\frac{1}{\hat{\lambda}} - \gamma}\right] & \text{si } \gamma < \frac{1}{\hat{\lambda}} \\ \infty & \text{si } \gamma \geq \frac{1}{\hat{\lambda}} \end{cases}
\end{aligned}$$

Utilizando el Lema 4, podemos afirmar que

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty e^{n\gamma\lambda} \pi(\mathrm{d}\lambda|T_1, \ldots, T_n) = \Lambda_{\hat{\lambda}}(\gamma).$$

Entonces, por el Teorema de *Gärtner Ellis*. Sabemos que $\pi(\cdot|T_1, \ldots, T_n)$ satisface un **LDP** con función tasa buena $J^{(1)}_{\hat{\lambda}} = \Lambda^*_{\hat{\lambda}}$ definida por

$$\Lambda^*_{\hat{\lambda}}(\lambda) := \sup_{\lambda\in\mathbb{R}}\{\gamma\lambda - \Lambda_{\hat{\lambda}}(\gamma)\} = \begin{cases} H(G_{1,\hat{\lambda}}|G_{1,\lambda}) & \text{si } \lambda > 0 \\ \infty & \text{si } \lambda \leq 0 \end{cases}$$

donde la última igualdad se sigue directamente de (40) con $\alpha = 1$ y $\beta = \frac{1}{\hat{\lambda}}$.

◇

Para evaluar las probabilidades de cruce de nivel predictivas de grandes desviaciones, necesitaremos primero establecer el siguiente resultado.

Lema 6. *Las función* $(\lambda, \ell) \longmapsto w(\lambda, \ell)$ *es continua.*

Demostración.

Para probar el Lema, demostraremos que para cualquier pareja $(\lambda, \ell) = (0, \infty) \times \mathcal{M}_1([0, M])$, el $\lim_{n\to\infty} w(\lambda_n, \ell_n) = w(\lambda, \ell)$ para cualquier sucesión $\{(\lambda_n, \ell_n)\}_{n\geq 1}$ tal que $\lim_{n\to\infty}(\lambda_n, \ell_n) = (\lambda, \ell)$. Por conveniencia, escribiremos simplemente w_n para denotar $w(\lambda_n, \ell_n)$ para $n \geq 1$. Consideraremos las funciones $\mathbb{H}(\gamma)$ y $\{\mathbb{H}_n(\gamma)\}_{n\geq 1}$ definidas como sigue

$$\mathbb{H}(\gamma) = \lambda(M_\ell(\gamma) - 1) - c\gamma \qquad \text{y} \qquad \mathbb{H}_n(\gamma) = \lambda_n(M_{\ell_n}(\gamma) - 1) - c\gamma.$$

Como $\{\ell_n\}_{n\geq} \subset \mathcal{M}_1([0, M])$ y $[0, M]$ es un conjunto compacto, entonces para cualquier $\gamma \in \mathbb{R}$ se cumple que $\lim_{n\to\infty} M_{\ell_n}(\gamma) = M_\ell(\gamma)$ y en consecuencia $\lim_{n\to\infty} \mathbb{H}_n(\gamma) =$

$\mathbb{H}(\gamma)$. De hecho $\mathbb{H}_n(w_n) = 0$ para cualquier $n \geq 1$, y es claro que cada función $\mathbb{H}_n$ es continua, supone valores finitos[7] y $\mathbb{H}_n(0) = 0$.

Dada la definición de $w(\lambda, \ell)$ antes expuesta, podemos decir que existen $a, b \in \mathbb{R}$ tales que $a < w(\lambda, \ell) < b$, donde a y b no dependen de λ y ℓ, pero cumplen que $\mathbb{H}(a) < 0$ y $\mathbb{H}(b) > 0$. Por lo tanto, existe $N \geq 1$ tal que $a \leq w_n \leq b$ para toda $n \geq N$. De hecho existe una subsucesión $\{w_{n_k}\}_{k\geq 1}$, de $\{w_n\}_{n\geq 1}$ tal que el $\lim_{k\to\infty} w_{n_k} = \eta$ para alguna $\eta \in [a, b]$ (Teorema de *Bolzano-Waiestrass*).

Si $\eta < w(\lambda, \ell)$ entonces existe $\eta_- \in (\eta, w(\lambda, \ell))$ tal que $w_{n_k} < \eta_-$ eventualmente y con esto tenemos que

$$0 = \mathbb{H}_{n_k}(w_{n_k}) < \mathbb{H}_{n_k}(\eta_-) \xrightarrow{k\to\infty} \mathbb{H}(\eta_-) < \mathbb{H}(w(\lambda, \ell)) = 0,$$

lo que es una contradicción.

De forma similar si $\eta > w(\lambda, \ell)$, existe $\eta_+ \in (w(\lambda, \ell), \eta)$ tal que $w_{n_k} > \eta_+$ eventualmente, y entonces

$$0 = \mathbb{H}_{n_k}(w_{n_k}) > \mathbb{H}_{n_k}(\eta_+) \xrightarrow{k\to\infty} \mathbb{H}(\eta_+) > \mathbb{H}(w(\lambda, \ell)) = 0,$$

lo cual es imosible. Por lo tanto, concluimos que cualquier subsucesión de $\{w_n\}_{n\geq 1}$ converge a $w(\lambda, \ell)$, con lo que termina la prueba. De hecho, si $\lim_{n\to\infty} w_n = w(\lambda, \ell)$ no se cumple, entonces existe $\epsilon_0 > 0$ y una subsucesión $\{w_{n_k}\}_{k\geq 1}$ de $\{w_n\}_{n\geq 1}$ tal que $|w_{n_k} - w(\lambda, \ell)| \geq \epsilon_0$. Tal subsucesión tiene a su vez una subsucesión convergente, pues $\{w_{n_k}\} \subset [\alpha, \beta]$ (Teorema de *Bolzano-Waiestrass*) y entonces $\{w_{n_k}\}$ converge a $w(\lambda, \ell)$ por el mismo razonamiento expuesto arriba y de aquí se tiene nuevamente una contradicción, por lo tanto $\lim_{n\to\infty} w_n = w(\lambda, \ell)$.

$\diamond$

Proposición 6. *Sean $k \geq 1$, $p_1, \ldots, p_k \geq 0$ tales que $\sum_{i=1}^k p_i = 1$ y consideremos la distribución apriori $\pi \otimes \mathscr{D}_\mu$ sobre (λ, ℓ) como en el Lema anterior. Supongamos también que $\bar{T}_n \to \frac{1}{\hat{\lambda}}$ y $\bar{\ell}_n \to \hat{\ell} \in supp(\mathscr{D}_\mu)$ conforme $n \to \infty$. Entonces para toda $q > 0$, tenemos*

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\{\tau_{qn}(\lambda, \ell) < \infty\} \pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) = -[\bar{w}(\hat{\lambda}, \hat{\ell})](q),$$

y

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\} \pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) = -[\bar{w}(\hat{\lambda}, \hat{\ell})](q),$$

donde

$$-[\bar{w}(\hat{\lambda}, \hat{\ell})](q) = \sup_{\lambda>0, \ell\in supp(\mathscr{D}_\mu)} \{-qw(\lambda, \ell) - [H(G_{1,\hat{\lambda}}|G_{1,\lambda}) + H(\hat{\ell}|\ell)]\}.$$

[7] $M_\ell(\gamma) < \infty$ por hipótesis.

Demostración.

Al principio de ésta sección establecimos que para toda $\epsilon > 0$, existe $n_\epsilon \geq 1$ tal que (45) se cumple para toda $n \geq n_\epsilon$. Entonces, por el Lema 5 y el Lema de *Varadhan* aplicado a la función continua $(\lambda, \ell) \longmapsto w(\lambda, \ell)$ obtenemos que

$$\begin{aligned}
&-[\bar{w}(\hat{\lambda}, \hat{\ell})](q+\epsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\{\tau_{qn}(\lambda,\ell) < \infty\}\pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\{\tau_{qn}(\lambda,\ell) < \infty\}\pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) \\
&\leq -[\bar{w}(\hat{\lambda}, \hat{\ell})](q)
\end{aligned}$$

y

$$\begin{aligned}
&-[\bar{w}(\hat{\lambda}, \hat{\ell})](q+\epsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}\pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \ln \int_{(0,\infty)\times\mathcal{M}_1([0,M])} \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}\pi \otimes \mathscr{D}_\mu(\mathrm{d}\lambda, \mathrm{d}\ell|[datos]_n) \\
&\leq -[\bar{w}(\hat{\lambda}, \hat{\ell})](q)
\end{aligned}$$

Para conconcluir la prueba, basta con argumentar que $-[\bar{w}(\hat{\lambda}, \hat{\ell})](q)$ es una función continua de $q > 0$. Este hecho se sigue facilmente al darnos cuenta de que también es una función convexa, propiedad heredada de la entroía relativa. De hecho es una función que toma valores finitos ya que $0 \geq -[\bar{w}(\hat{\lambda}, \hat{\ell})](q) \geq -qw(\hat{\lambda}, \hat{\ell})$ para toda $q > 0$ y por lo tanto la continuidad de $-[\bar{w}(\hat{\lambda}, \hat{\ell})](q)$ se sigue del hecho de que una función convexa en un intervalo abierto, es continua en ese mismo intervalo (ver [Rud86]).

◇

5.4 Distribución ℓ Concida y Concentrada en $[0, M]$, con $M > 0$.

En ésta sección se presentan resultados para la probabilidad predictiva de cruce de nivel, en el caso particular en que ℓ es conocida y concentrada en $[0, M]$, tomaremos a λ desconocida y a la muestra de tamaño n, $[datos]_n = (T_1, \ldots, T_n)$. Supondremos adicionalmente que π es la distribución apriori de λ, tal y como en la sección pasada.

Adaptando la prueba de la Proposición 6 y considerando la función (que depende de $q > 0$)

$$-[\bar{w}^\ell(\hat{\lambda})](q) = \sup_{\lambda>0}\{-qw(\lambda,\ell) - H(G_{1,\hat{\lambda}}|G_{1,\lambda})\}$$

Afirmamos que

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda,\ell) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n) = -[\bar{w}^\ell(\hat{\lambda})](q)$$

y

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}\pi(\mathrm{d}\lambda|[datos]_n) = -[\bar{w}^\ell(\hat{\lambda})](q),$$

pues bajo el mismo argumento expuesto para demostrar la Proposición 6, sabemos que para toda $\epsilon > 0$ existe $n_\epsilon \geq 1$ tal que (45) se cumple para toda $n \geq n_\epsilon$. Por lo tanto aplicando el Lema 5 junto con el Lema de *Varadhan* a la función continua $\lambda \mapsto w(\lambda)$ con

$$w(\lambda) := \sup\{\gamma \geq 0 \ : \ \lambda(M_\ell(\gamma) - 1) - c\gamma \leq 0\},$$

obtenemos que

$$\begin{aligned}
&- [\bar{w}^\ell(\hat{\lambda})](q + \epsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda, \ell) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda, \ell) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n) \\
&\leq -[\bar{w}^\ell(\hat{\lambda})](q)
\end{aligned}$$

y

$$\begin{aligned}
&- [\bar{w}^\ell(\hat{\lambda})](q + \epsilon) \\
&\leq \liminf_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}\pi(\mathrm{d}\lambda|[datos]_n) \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\Big\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\Big\}\pi(\mathrm{d}\lambda|[datos]_n) \\
&\leq -[\bar{w}^\ell(\hat{\lambda})](q),
\end{aligned}$$

concluyendo además que $-[\bar{w}^\ell(\hat{\lambda})](q)$ es continua bajo el mismo argumento de convexidad de la sección anterior [Rud86].

◇

Por otro lado, si en lugar de los datos, $[datos]_n$, tomamos en cuenta la muestra de tamaño n

$$[datos\ alternativos]_n = [N_1, N_2 - N_1, \ldots, N_n - N_{n-1}),$$

(como en [Mac04]), donde $N_t = \sum_{k\geq 1} \mathbb{I}_{\{T_1+\cdots+T_k \leq t\}}$. Considerando ahora a la función (de $q > 0$)

$$-[\bar{w}_\ell(\hat{\lambda})](q) = \sup_{\lambda>0}\{-qw(\lambda, \ell) - H(P_{\hat{\lambda}}|P_\lambda)\},$$

donde $H(P_{\hat{\lambda}}|P_\lambda)$ es la entropía relativa entre dos distribuciones Poisson $P_{\hat{\lambda}}$ y P_λ presentada al inicio de la sección 5.

Entonces podemos extender la prueba de la *Proposición 3* de [Mac04], donde la ley ℓ coincide con la ley δ_1 de la variable aleatoria constante igual a 1. Aquí ℓ está concentrada en un conjunto acotado y la distribución apriori de λ es una mezcla finita de distribuciones gama $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$. De este modo

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda, \ell) < \infty\}\pi(\mathrm{d}\lambda|[datos\ alternativos]_n) = -[\bar{w}_\ell(\hat{\lambda})](q)$$

y

$$\lim_{n\to\infty}\frac{1}{n}\ln\int_0^\infty \mathbb{P}\Big\{\tau_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big\}\pi(\mathrm{d}\lambda|[datos\ alternativos]_n)=-[\bar{w}_\ell(\hat{\lambda})](q),$$

conforme $\frac{N_n}{n}$ tiende a $\hat{\lambda}$ cuando n tiende a ∞.

Finalmente es interesante comparar $-[\bar{w}_\ell(\hat{\lambda})](q)$ con $-[\bar{w}^\ell(\hat{\lambda})](q)$. Para este fin, hagamos uso de (39) para darnos cuenta de que

$$-[\bar{w}_\ell(\hat{\lambda})](q)=\sup_{\lambda>0}\{-qw(\lambda,\ell)-\hat{\lambda}H(G_{1,\hat{\lambda}}|G_{1,\lambda})\},$$

de modo que $-[\bar{w}_\ell(\hat{\lambda})](q)$ y $-[\bar{w}^\ell(\hat{\lambda})](q)$ coinciden cuando $\hat{\lambda}=1$. De hecho, $-[\bar{w}_\ell(\hat{\lambda})](q)\le -[\bar{w}^\ell(\hat{\lambda})](q)$ para $\hat{\lambda}\ge 1$ y $-[\bar{w}_\ell(\hat{\lambda})](q)\ge -[\bar{w}^\ell(\hat{\lambda})](q)$ para $\hat{\lambda}\le 1$.

5.5 Distribución $exp(\cdot|\beta)$.

Supongamos que $\ell \overset{D}{=} exp(\cdot|\beta)$, es decir que ℓ es la distribución exponencial con tasa de fallo igual a β. En éste caso los saltos no son acotados por una constante positiva M y por lo tanto no podemos adaptar las técnicas presentadas en la prueba de la Proposición 6.

De manera más precisa, si $(\pi(\cdot|[datos]_n)$ satisface un **LDP**, de la desigualdad de *Lunberg* (43) podemos derivar cotas superiores para

$$\limsup_{n\to\infty}\frac{1}{n}\ln\int_{(0,\infty)\times\mathcal{M}_1([0,\infty))}\mathbb{P}\{\tau_{qn}(\lambda,\ell)<\infty\}\pi(\mathrm{d}\lambda,\mathrm{d}\ell|[datos]_n),$$

y

$$\limsup_{n\to\infty}\frac{1}{n}\ln\int_{(0,\infty)\times\mathcal{M}_1([0,\infty))}\mathbb{P}\Big\{\tau_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big\}\pi(\mathrm{d}\lambda,\mathrm{d}\ell|[datos]_n)$$

vía el Lema de *Varadhan* y el *Principio de Contracción*. Sin embargo, no podemos derivar cotas inferiores para

$$\liminf_{n\to\infty}\frac{1}{n}\ln\int_{(0,\infty)\times\mathcal{M}_1([0,\infty))}\mathbb{P}\{\tau_{qn}(\lambda,\ell)<\infty\}\pi(\mathrm{d}\lambda,\mathrm{d}\ell|[datos]_n),$$

y

$$\liminf_{n\to\infty}\frac{1}{n}\ln\int_{(0,\infty)\times\mathcal{M}_1([0,\infty))}\mathbb{P}\Big\{\tau_q\Big(n\lambda,\frac{1}{n}\ell\Big)<\infty\Big\}\pi(\mathrm{d}\lambda,\mathrm{d}\ell|[datos]_n).$$

Recordemos que (46) provee una expresión en forma cerrada para la probabilidad de ruina $\mathbb{P}\{\tau_Q(\lambda,\ell)<\infty\}$ cuando $\ell \overset{D}{=} exp(\cdot|\beta)$, por otro lado, cuando no es posible encontrar una constante $\tilde{M}>0$ tal que

$$1-\frac{w(\lambda,\exp(\cdot|\beta))}{\beta}\ge\tilde{M},\qquad\forall\ (\lambda,\beta),$$

en un conjunto de probabilidad 1 con respecto a la distribución apriori, las cotas inferiores no pueden encontrarse.

De momento, la elección (conjugada) natural de productos de distribuciones gama en (λ,β), tiene soporte en $(0,\infty)\times(0,\infty)$ pero no resuleve el problema, pues

$$\lim_{\beta\to\infty} \frac{w(\lambda, \exp(\cdot|\beta))}{\beta} = 1,$$

para cada $\lambda > 0$ fija, y

$$\lim_{\lambda\to\infty} \frac{w(\lambda, \exp(\cdot|\beta))}{\beta} = 1,$$

para cada $\beta > 0$ fija.

Una manera de abordar el problema, es suponer que $\beta = k\lambda$ para algún $k > 0$, de tal forma que

$$1 - \frac{w(\lambda, \exp(\cdot|\beta))}{\beta} = 1 - \frac{w(\lambda, \exp(\cdot|k\lambda))}{k\lambda} = \text{mín}\left\{\frac{1}{kc}, 1\right\} > 0,$$

que no depende de λ. Entonces la desigualdad

$$1 - \frac{w(\lambda, \exp(\cdot|\beta))}{\beta} \geq \tilde{M},$$

se cumple, con la igualdad cuando $\tilde{M} = \text{mín}\left\{\frac{1}{kc}, 1\right\}$.

Si suponemos que $k > 0$ es conocida, λ se convierte en el único parámetro desconocido. Considerando la muestra de tamaño n, dada por

$$[datos]_n = ((B_1, T_1), \ldots, (B_n, T_n)),$$

implica que la verosimilitud es

$$L(\lambda|[datos]_n) = k\lambda^{-k\lambda B_1}\lambda e^{-\lambda T_n} \cdots k\lambda^{-k\lambda B_n}\lambda e^{-\lambda T_n} = k^n \lambda^{2n} e^{\lambda n[k\bar{B}_n + \bar{T}_n]},$$

donde $\bar{B}_n = \frac{1}{n}\sum_{k=1}^n B_k$ y $\bar{T}_n = \frac{1}{n}\sum_{k=1}^n T_k$ son las medias empíricas de $\{B_1, \ldots, B_n\}$ y $\{T_1, \ldots, T_n\}$ respectivamente. Por lo tanto, $k\bar{B}_n + \bar{T}_n$ es una estadística suficiente.

Con la distribución apriori conjugada para λ, $G_{\alpha,\beta}$, tenemos que la distribución posterior es

$$G_{\alpha,\beta}(\cdot|[datos]_n) = G_{\alpha_n,\beta_n},$$

donde $\alpha_n = \alpha + 2n$ y $\beta_n = \beta + n(k\bar{B}_n + \bar{T}_n)$.

A continuación probaremos un LDP para las distribuciones posteriores.

Lema 7. *Sean $k \geq 1$, $p_1, \ldots, p_n \geq 0$ tales que $\sum_{i=1}^k p_i = 1$ y consideremos la distribución apriori $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$ para λ. Supongamos que $k\bar{B}_n + \bar{T}_n \longrightarrow \frac{2}{\hat{\lambda}}$ conforme $n \to \infty$. Entonces $\pi(\cdot|[datos]_n)$ satisface un principio de gandes desviaciones con función tasa buena $I_{\hat{\lambda}}$ definida por*

$$I^*_{\hat{\lambda}}(\lambda) = \sup_{\gamma\in\mathbb{R}}\{\gamma\lambda - \Lambda_{\hat{\lambda}}(\gamma)\} = \begin{cases} H(G_{2,\hat{\lambda}}|G_{2,\lambda}) & si\ \lambda > 0 \\ \infty & si\ \lambda \leq 0 \end{cases}$$

Demostración.

Sea $\gamma \in \mathbb{R}$ arbitrario y fijo. Como $k\bar{B}_n + \bar{T}_n \to \frac{2}{\hat{\lambda}}$ conforme $n \to \infty$, tenemos que para cualquier distribución apriori conjugada $G_{\alpha,\beta}$

$$\begin{aligned}
\Lambda_{\hat{\lambda}}(\gamma) &= \lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty e^{n\gamma\lambda} G_{\alpha,\beta}(\mathrm{d}\lambda|[datos]_n) \\
&= \lim_{n\to\infty} \begin{cases} \frac{\alpha+2n}{n} \ln\left(\frac{\beta+n(k\bar{B}_n+\bar{T}_n)}{\beta+n(k\bar{B}_n+\bar{T}_n)-n\gamma}\right) & \text{si } n\gamma < \beta + n(k\bar{B}_n + \bar{T}_n) \\ \infty & \text{si } n\gamma \geq \beta + n(k\bar{B}_n + \bar{T}_n) \end{cases} \\
&= \begin{cases} 2\left(\frac{\frac{2}{\hat{\lambda}}}{\frac{2}{\hat{\lambda}}-\gamma}\right) & \text{si } \gamma < \frac{2}{\hat{\lambda}} \\ \infty & \text{si } \gamma \geq \frac{2}{\hat{\lambda}} \end{cases}
\end{aligned}$$

de hecho como

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty e^{n\gamma\lambda} \pi(\mathrm{d}\lambda|[datos]_n) = \Lambda_{\hat{\lambda}}(\gamma)$$

se cumple por el Lema 4. Entonces utilizando el Teorema de *Gärtner Ellis*, sabemos que $\pi(\cdot|[datos]_n)$ satisface un **LDP** con función tasa buena $I_{\hat{\lambda}} = \Lambda^*_{\hat{\lambda}}$ definida por

$$\Lambda^*_{\hat{\lambda}}(\lambda) := \sup_{\gamma\in\mathbb{R}} \{\gamma\lambda - \Lambda_{\hat{\lambda}}(\gamma)\} = \begin{cases} H(G_{2,\hat{\lambda}}|G_{2,\lambda}) & \text{si } \lambda > 0 \\ \infty & \text{si } \lambda \leq 0 \end{cases}$$

donde la última igualdad se sigue de (40) con $\alpha = 2$ y $\beta = \frac{2}{\hat{\lambda}}$.

◇

Para concluir esta sección, estimaremos las probabilidades de cruce de nivel para las cuales, la siguiente función (de $q > 0$) es útil.

$$-[\bar{w}(\hat{\lambda}, \exp(\cdot|k\hat{\lambda}))](q) = \sup_{\lambda>0} \{-qw(\lambda, \exp(\cdot|k\lambda)) - H(G_{2,\hat{\lambda}}|G_{2,\lambda})\}.$$

Proposición 7. *Sean $k \geq 1$, $p_1, \ldots, p_n \geq 0$ tales que $\sum_{i=1}^k p_i = 1$ y consideremos la distribución apriori $\pi = \sum_{i=1}^k p_i G_{\alpha(i),\beta(i)}$ para λ. Supongamos que $k\bar{B}_n + \bar{T}_n \longrightarrow \frac{2}{\hat{\lambda}}$ conforme $n \to \infty$. Entonces, para toda $q > 0$ tenemos que*

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda, \exp(\cdot|k\lambda)) < \infty\} \pi(\mathrm{d}\lambda|[datos]_n) = -[\bar{w}(\hat{\lambda}, \exp(\cdot|k\hat{\lambda}))](q),$$

y

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_0^\infty \mathbb{P}\{\tau_q(n\lambda, \exp(\cdot|nk\lambda)) < \infty\} \pi(\mathrm{d}\lambda|[datos]_n) = -[\bar{w}(\hat{\lambda}, \exp(\cdot|k\hat{\lambda}))](q).$$

Demostración.

Fijemos $\tilde{M} = \min\left\{\frac{1}{kc}, 1\right\}$ y sea $n \geq 1$ fija y arbitraria. Entonces por (46), tenemos que

$$\tilde{M} e^{-w(\lambda,\exp(\cdot|k\lambda))qn} \leq \mathbb{P}\{\tau_{qn}(\lambda, \exp(\cdot|k\lambda)) < \infty\} \leq e^{-w(\lambda,\exp(\cdot|k\lambda))qn}$$

y

$$\tilde{M}e^{-w(\lambda,\exp(\cdot|k\lambda))qn} \leq \mathbb{P}\{\tau_q(n\lambda,\exp(\cdot|nk\lambda)) < \infty\} \leq e^{-w(\lambda,\exp(\cdot|k\lambda))qn},$$

en donde para estas últimas desigualdades utilizamos que $w(n\lambda,\exp(\cdot|nk\lambda)) = nw(\lambda,\exp(\cdot|k\lambda))$.

Por lo tanto

$$\begin{aligned}
&-[\bar{w}(\hat{\lambda},\exp(\cdot|k\hat{\lambda}))](q)\\
&\leq \liminf_{n\to\infty}\frac{1}{n}\ln\int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda,\exp(\cdot|k\lambda)) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n)\\
&\leq \limsup_{n\to\infty}\frac{1}{n}\ln\int_0^\infty \mathbb{P}\{\tau_{qn}(\lambda,\exp(\cdot|k\lambda)) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n)\\
&\leq -[\bar{w}(\hat{\lambda},\exp(\cdot|k\hat{\lambda}))](q)
\end{aligned}$$

y

$$\begin{aligned}
&-[\bar{w}(\hat{\lambda},\exp(\cdot|k\hat{\lambda}))](q)\\
&\leq \liminf_{n\to\infty}\frac{1}{n}\ln\int_0^\infty \mathbb{P}\{\tau_q(n\lambda,\exp(\cdot|nk\lambda)) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n)\\
&\leq \limsup_{n\to\infty}\frac{1}{n}\ln\int_0^\infty \mathbb{P}\{\tau_n(n\lambda,\exp(\cdot|nk\lambda)) < \infty\}\pi(\mathrm{d}\lambda|[datos]_n)\\
&\leq -[\bar{w}(\hat{\lambda},\exp(\cdot|k\hat{\lambda}))](q),
\end{aligned}$$

por el Lema anterior y el Lema de *Varhadan* aplicado a la función continua $\lambda \longmapsto w(\lambda,\exp(\cdot|k\lambda))$.

◇

6 Algunas consideraciones sobre el enfoque Frecuentista y el enfoque Bayesiano.

A continuación mencionaremos algunas características que distinguen el enfoque *Bayesiano* del *Frecuentista* para los resultados en donde se involucran un **LDP** y estimaciones de probabilidades de cruce de nivel.

Una primera observación que es importante hacer, es que a partir del **LDP** las funciones tasa para las distribuciones posteriores, se pueden expresar en términos de la misma entropía relativa $H(\cdot|\cdot)$ utilizada para la función tasa correspondiente a una sucesión adecuada de estadísticas suficientes $(T_n([datos]_n))$. No obstante, en este último caso, el papel jugado por los argumentos de $H(\cdot|\cdot)$ es intercambiado. Este hecho puede explicarse dada la diferencia entre los papeles adoptados por el parámetro θ y los datos, en las perspectivas *Bayesiana* y *Frecuentista* con respecto a un modelo estadístico $(X,\mathcal{B}_X,(F(\theta):\theta\in\Theta))$. De hecho, para el **LDP** de la estadística suficiente, nos preguntamos cuan probable es para la estadística suficiente estar cerca de un valor $\hat{\theta}$ cuando el verdadero valor del parámetro es θ y la función tasa es de la forma $\hat{\theta}\mapsto I_\theta(\hat{\theta}) = H(F(\hat{\theta})|F(\theta))$. Mientras que por otro lado, en el **LDP** de las distribuciones posteriores, nos preguntamos cuan probable es para el valor verdadero del parámetro estar cerca de θ dado que observamos la estadística suficiente cerca de algún $\hat{\theta}$ y la función tasa es de la forma $\theta \mapsto I_{\hat{\theta}}(\theta) = H(F(\hat{\theta})|F(\theta))$.

Sean C° y $\bar{C}$, el interior y la cerradura de C respectivamente. Entonces

$$\lim_{n\to\infty} \mathbb{P}_\theta\{T_n([datos]) \in C\} = 0, \qquad \text{si} \quad \theta \notin \bar{C},$$

donde $\mathbb{P}_\theta$ es la medida de probabilidad correspondiente a la función de distribución $F(\theta)$ y

$$\lim_{n\to\infty} \pi(C|[datos_n]) = 0 \qquad \text{si} \quad \hat{\theta} \notin \bar{C}.$$

Por el **LDP** de la estadística suficiente y el **LDP** de las distribuciones posteriores, podemos decir que $\mathbb{P}_\theta\{T_n([datos]) \in C\}$ y $\pi(C|[datos_n])$ decrecen a cero exponencialmente conforme n tiende a infinito, bajo las siguientes hipótesis.

- Supongamos que $I_\theta(C) = \inf_{\hat{\theta}\in C} I_\theta(\hat{\theta})$ y $I_\theta(C) = I_\theta(C^{\circ}) = I_\theta(\bar{C}) \in (0,\infty)$.[8] Entonces, para toda $\epsilon > 0$, existe $n_\epsilon \geq 1$ tal que para toda $n \geq n_\epsilon$.

$$e^{-n(I_\theta(C)+\epsilon)} < \mathbb{P}_\theta\{\tau_n([datos]_n) \in C\} < e^{-n(I_\theta(C)-\epsilon)}.$$

- Supongamos que $I_{\hat{\theta}}(C) = \inf_{\theta\in C} I_{\hat{\theta}}(\theta)$ y $I_{\hat{\theta}}(C) = I_{\hat{\theta}}(C^{\circ}) = I_{\hat{\theta}}(\bar{C}) \in (0,\infty)$.[9] Entonces, para toda $\epsilon > 0$, existe $n_\epsilon \geq 1$ tal que para toda $n \geq n_\epsilon$.

$$e^{-n(I_{\hat{\theta}}(C)+\epsilon)} < \pi(C|[datos]_n) < e^{-n(I_{\hat{\theta}}(C)-\epsilon)}.$$

La segunda diferencia, está relacionada con las estimaciones de probabilidades de cruce de nivel. Para ver esto, es importante recordar los siguientes límites conforme n tiende a infinito necesarios para describir las estimaciones *Frecuentistas* de las probabilidades de cruce de nivel.

Sea $q > 0$, entonces

- Para el proceso Poisson Compuesto con saltos hacia arriba y deriva negativa, con ℓ concentrada en $[0, M]$ para $M > 0$ tenemos

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\{\tau_{qn}(\lambda,\ell) < \infty\} = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\{\tau_q\Big(n\lambda, \frac{1}{n}\ell\Big) < \infty\} = -qw(\lambda,\ell),$$

  por (45), ya que es $\epsilon > 0$ es arbitrario.

- Para el proceso Poisson Compuesto con saltos hacia arriba y deriva negativa, con ℓ igual a la distribución exponencial de tasa β tenemos

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\{\tau_{qn}(\lambda, \exp(\cdot|\beta)) < \infty\} = \lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\{\tau_q\Big(n\lambda, \exp(\cdot|n\beta)\Big) < \infty\} = -qw(\lambda, \exp(\cdot|\beta)),$$

  por (46), donde para el segundo límite utilizamos la identidad

$$w(n\lambda, \exp(\cdot|nk\lambda)) = nw(n\lambda, \exp(\cdot|nk\lambda)).$$

[8]Notemos que $I_\theta(C) > 0$ pues $\theta \notin \bar{C}$.

[9]Notemos que $I_{\hat{\theta}}(C) > 0$ pues $\hat{\theta} \notin \bar{C}$.

Es importante notar que en todos los casos considerados, los límites tienen la misma forma, $-qw(\theta)$, con

$$w(\theta) = \sup\{\gamma \geq 0 : \ln \mathbb{E}[e^{\gamma Z_1^\theta}] \leq 0\},$$

siendo una función lineal de $q > 0$ para cada valor posible fijo del parámetro $\theta \in \Theta$.

Las estimaciones frecuentistas de los límites expuestos arriba, son $-qw(\hat{\theta})$ donde es el límite casi seguro de la estadística suficiente. En el contexto Bayesiano considerado en la secciones anteriores, las estimaciones son

$$-[\bar{w}(\hat{\theta})](q) = \sup_{\theta \in \Theta}\{-qw(\theta) - H(F(\hat{\theta})|F(\theta))\}, \tag{47}$$

donde $(X, \mathcal{B}_X, (F(\theta) : \theta \in \Theta))$ es el modelo estadístico expuesto arriba.

Al comparar $-[\bar{w}(\hat{\theta})](q)$ con $-qw(\hat{\theta})$, podemos darnos cuenta facilmente por (47), que la diferencia $-[\bar{w}(\hat{\theta})](q) - (-qw(\hat{\theta}))$ es no negativa y utilizando la proposición 8 del apéndice B podemos afirmar que es una función no decreciente de $q > 0$. Estas consideraciones señalan que las estimaciones Bayesianas de probabilidades de cruce de nivel expuestas en este trabajo son más conservativas a un grado que se vuelve más pronuciado conforme q crece.

Agradecimientos. *Quiero agradecer a la* Dra. Ana Meda *por todas sus enseñanzas, haber dirigido este trabajo y haber brindando su tiempo y enorme paciencia para la revisión del mismo. Agradezco también al* Dr. Raúl Rueda *por haberme introducido a la* Estadística Bayesiana no Paramétrica *y orientarme sobre el tema. Y por supuesto, a la UNAM.*

7 Apendice A.

En esta sección enunciaremos algunos resultados clásicos de la teoría de Grandes Desviaciones utilizados a lo largo de este documento. Las demostraciones de todos los resultados enunciados aquí, serán referidas al libro de *A. Dembo y O. Zeitouni* [DZ93].

Teorema 6. Principio de Contracción. *Sean $\mathcal{X}$ y $\mathcal{Y}$ dos espacios topológicos de Hausdorff y $f : \mathcal{X} \longmapsto \mathcal{Y}$ una función continua. Consideremos una función tasa buena $I : \mathcal{X} \longmapsto [0, \infty]$.*

(a) Para cada $y \in \mathcal{Y}$, si definimos

$$I'(y) := \inf\{I(x) : \ x \in \mathcal{X}, \ y = f(x)\}.$$

Entonces I' es una función tasa buena en $\mathcal{Y}$. [10]

*(b) Si I controla el **LDP** asociado con una familia de medidas de probabilidad $\{\mu_n\}$ en $\mathcal{X}$, entonces I' controla el **LDP** asociado con la familia de medidas de probabilidad $\{\mu_n \circ f^{-1}\}$ en $\mathcal{Y}$.*

Definición 7. *Sea $\Lambda : \mathbb{R}^m \longmapsto (-\infty, \infty]$ una función convexa y sea*

$$\mathcal{D}_\Lambda = \{\gamma \in \mathbb{R}^m : \Lambda(\gamma) < \infty\},$$

con interior $\mathcal{D}_\Lambda^\circ$. Entonces, Λ es una función esencialmente suave si:

(i) $\mathcal{D}_\Lambda^\circ$ es no vacío.

(ii) Λ es diferenciable en todo $\mathcal{D}_\Lambda^\circ$ y es tal que $\lim_{n\to\infty} |\nabla\Lambda(\gamma_n)| = \infty$, siempre que γ_n en $\mathcal{D}_\Lambda^\circ$ converga a un punto frontera de $\mathcal{D}_\Lambda^\circ$.

Sea Ω un espacio topológico de *Hausdorff* con σ-álgebra de *Borel* $\mathcal{B}_\Omega$ y sea $\{\nu_n\}$ una sucesión de medidas de probabilidad en $(\Omega, \mathcal{B}_\Omega)$.

Teorema 7. Gärtner – Ellis. *Consideremos $(\Omega, \mathcal{B}_\Omega) = (\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ y supongamos que para toda $\gamma \in \mathbb{R}^m$, el límite*

$$\Lambda(\gamma) = \lim_{n\to\infty} \frac{1}{n} \ln \int_{\mathbb{R}^m} e^{n\gamma'\boldsymbol{x}} \nu_n(\mathrm{d}\boldsymbol{x}),$$

*existe como un real extendido. Si Λ es una función esencialmente suave y semicontinua por abajo, entonces se cumple el **LDP** con función tasa buena $\Lambda^* : \mathbb{R}^m \longmapsto (-\infty, \infty]$ definida por*

$$\Lambda^*(\boldsymbol{x}) = \sup_{\gamma\in\mathbb{R}^m} \{\gamma'\boldsymbol{x} - \Lambda(\gamma)\}.$$

[10] Como siempre, el ínfimo sobre un conjunto vacío se define como infinito.

Lema 8. Lema de Varadhan. *Supongamos que* $\{\nu_n\}$ *satisface el* ***LDP*** *con función tasa buena* $I : \Omega \longmapsto [0, \infty]$ *y sea* $\phi : \Omega \mapsto \mathbb{R}$ *una función continua. Supongamos además cualquiera de las siguientes dos condiciones*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln \int_\Omega e^{n\phi(w)} \mathbb{I}_{\{\phi(w)\geq M\}} \nu_n(\mathrm{d}w) = -\infty,$$

o la condición de momentos.

$$\limsup_{n\to\infty} \frac{1}{n} \ln \int_\Omega e^{\gamma n\phi(w)} \nu_n(\mathrm{d}w) < \infty,$$

para algún $\gamma > 1$*. Entonces*

$$\lim_{n\to\infty} \frac{1}{n} \ln \int_\Omega e^{n\phi(w)} \nu_n(\mathrm{d}w) = \sup_{w\in\Omega} \{\phi(w) - I(w)\}.$$

Sea Ω un espacio polaco, $\mathcal{M}_1(\Omega)$ al espacio de medidas de probabilidad en Ω, $\{X\}_{k\geq 1}$ una sucesión de variables aleatorias independientes con valores en Ω y ley común $\mu \in \mathcal{M}_1(\Omega)$. Denotemos con L_n a la medida empírica correspondiente a las primeras n observaciones

$$L_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k},$$

y sea $\mathcal{L}(L_n)$ la ley de L_n. Para $\nu \in \mathcal{M}_1(\Omega)$ se define la distancia de *Kullback-Leibler* o *Entropía Relativa* (relativa a μ) como

$$H(\nu|\mu) = \begin{cases} \int_\Omega \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \ln\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu & \nu \ll \mu \\ \infty & c.o.c \end{cases}$$

Teorema 8. Sanov. *La sucesión de medidas empíricas* L_n *satisface el* ***LDP*** *en* $\mathcal{M}_1(\Omega)$ *equipado con la topología* τ*, con función tasa buena convexa* $H(\cdot|\mu)$*.*

8 Apendice B.

Proposición 8. *Sea* $f : [0, \infty) \longmapsto \mathbb{R}$ *una función convexa, tal que*

$$0 \geq f(q) \geq -qm,$$

para toda $q \in [0, \infty)$ *y alguna* $m > 0$*. Entonces la función* g *definida por*

$$g(q) := f(q) - (-qm),$$

en no decreciente.

Demostración.

Consideremos $0 < x < y$ y fijemos $t = \frac{x}{y}$. Entonces

$$f(x) = f(ty + (1-t)0) \leq tf(y) + (1-t)f(0) = tf(y),$$

puesto que $x = ty + (1-t)0$, $t \in (0,1)$ y f es una función convexa. De hecho

$$(1-t)f(x) = f(x) - tf(x) \leq tf(y) - tf(x) = t(f(y) - f(x)),$$

de donde

$$f(y) - f(x) \geq \frac{1-t}{t} f(x) \geq \frac{1-t}{t}(-xm) = -\frac{x}{t}m + xm = -ym + xm,$$

que es equivalente a $g(y) \geq g(x)$.

$\diamond$

Referencias

[AN95] S. Asmussen and H.M. Nielsen. Ruin probabilities via local adjustment coefficients. *Journal of Applied Probability.*, 1995.

[Ant74] C.E. Antoniak. Mixtures of dirichlet processes with aplications to bayesian non prametric problems. *The Annals of Statistics.*, 1974.

[Asm87] S. Asmussen. *Applied Probability and Queues.* John Wiley and Sons, 1987.

[Asm00] S. Asmussen. *Ruin Probabilities.* World Scientific, 2000.

[BK54] Gnedenko. B.V. and A.N. Kolmogorov. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, 1954.

[BM73] T. Blackwell and J. MacQueen. Ferguson distributions via polya urn schemes. *The Annals of Statistics.*, 1973.

[DeG70] M.H. DeGroot. *Optimal Statistical Decisions.* Mc Graw-Hill., 1970.

[DH83] S. Dilal and W.J. Hall. Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society.*, 1983.

[DR97] P. Dupuis and Ellis R.S. *A weak Convergence Approach to the Theory of Large Deviations.* John Wiley and Sons., 1997.

[DY85] P. Diaconis and D. Ylvisaker. *Quantifying Prior Opinion.* In *Bayesian Statistics 2.*, J.M. Bernardo et al.,eds, 1985.

[DZ93] A. Dembo and O Zeitouni. *Large Deviations Techniques and Applications.* Jones and Bartlett, 1993.

[EKM97] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events.* Springer, 1997.

[Ess32] F. Esscher. On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift*, 1932.

[Fer73] T.S. Ferguson. A bayesian analysis of some nonparametric problems. *The Annals of Statistics.*, 1973.

[FK72] T.S. Ferguson and M.J. Klass. A representation of independent increment processes without gaussian components. *The Annals of Mathematical Statistics.*, 1972.

[GGOP98] A. Ganesh, P. Green, N. O'Connell, and S. Pitts. Bayesian network management. *Queueing Systems Theory Appl.*, 1998.

[GO99] A. Ganesh and N. O'Connell. An inverse of sanov's theorem. *Statist. Probab. Lett.*, 1999.

[GO00] A. Ganesh and N. O'Connell. A large deviation principle for dirichlet posteriors. *Bernoulli.*, 2000.

[Goo65] I.J. Good. The estimation of probabilities: An essay on modern bayesian methods. *Research Monograph No. 30, M.I.T. Press*, 1965.

[Gor88] H.O. Gorgii. *Gibbs Measures and Phase Transitions.* De Gruyter., 1988.

[Gra91] J. Grandell. *Aspects of Risk Theory.* Springer., 1991.

[Mac04] C. Macci. Large deviation estimates for some level crossing probabilities in bayesian setting. *Rend. Sem. Mat. Univ. Pol. Torino.*, 2004.

[Mac10] C. Macci. Large deviations for estimators of unknown probabilities, with applications in risk theory. *Statistics and Probability Letters.*, 2010.

[MP06] C. Macci and L. Petrella. Mixtures of conjugate prior distributions and large deviations for level crossing probabilities. *The Indian Journal of Statistics.*, 2006.

[OF04] A. O'Hagan and J.J. Foster. *Bayesian Inference. 2nd edition.* Arnold, 2004.

[Roc74] R.T. Rockafellar. *Conjugate Duality and Optimization.* SIAM., 1974.

[Rud86] W. Rudin. *Real and Complex Analysis. 3rd edition.* Mc Graw-Hill., 1986.

[Var08] S.R.S. Varadhan. Large deviations. *The Annals of Probability.*, 2008.

[Wil62] S. Wilks. *Mathematical Statistics.* John Wiley and Sons, 1962.