

***GAUSSIAN PROCESS REGRESIÓN COMO TÉCNICA BAYESIANA NO  
PARAMÉTRICA  
DOS APLICACIONES ACTUARIALES: VIDA Y NO VIDA***

***GAUSSIAN PROCESS REGRESSION AS A NON-PARAMETRIC BAYESIAN  
TECHNIQUE  
TWO ACTUARIAL APPLICATIONS: LIFE AND NON-LIFE***

David Rius Carretero<sup>1</sup>, Salvador Torra Porras

Agradecimientos:

Quería agradecer el amor, apoyo y toda la paciencia que han tenido mis padres, hermano y pareja durante estos años de formación académica y profesional, que no ha sido poca, y dedicarles este trabajo como el final de un ciclo, al igual que al resto de mi familia y amigos, por su apoyo incondicional.

---

<sup>1</sup> Autor para correspondencia: [davidrius96@hotmail.com](mailto:davidrius96@hotmail.com)

## RESUMEN

En este trabajo se ha realizado una breve introducción sobre que es una Regresión de Proceso Gaussiano (GPR) y dos aplicaciones que pueden integrarse en el ámbito de las Ciencias Actuariales. Por un lado, se ha realizado un ejercicio de interpolación sobre las tablas de mortalidad PASEM Unisex 2020, concluyendo que el GPR es una fuerte herramienta de interpolación y permite una tarificación más ajustada en el ramo de Vida. Por otro lado, se ha integrado el GPR como medida de predicción de provisiones en los ramos de No-Vida, obteniendo unos datos de salida prometedores. Por último, se concluye que un GPR puede ser un instrumento útil, siempre y cuando, se realice una buena selección del Kernel y un correcto período de entrenamiento del modelo.

Palabras Clave: Proceso Gaussiano, Normal Multivariante, Covarianza, Actuariales, Distribuciones.

## ABSTRACT

In this paper, a brief introduction has been made about what a Gaussian Process Regression (GPR) is and two applications that can be integrated in the field of Actuarial Sciences. On the one hand, an interpolation exercise has been carried out on the PASEM Unisex 2020 mortality tables, concluding that the GPR is a strong interpolation tool and allows a more adjusted pricing in the Life branch. On the other hand, the GPR has been integrated as a prediction measure for provisions in the Non-Life branches, obtaining promising output data. Finally, it is concluded that a GPR can be a useful instrument, as long as a good selection of the Kernel and the model training period are made.

KeyWords: Gaussian Process, Multivariate Normal, Covariance, Actuarial, Distributions.

## ÍNDICE

I	INTRODUCCIÓN .....	4
II	OBJETIVOS Y METODOLOGÍA .....	5
III	CAPITULO I – PROCESO GAUSSIANO .....	6
III.I	DISTRIBUCIÓN GAUSSIANA .....	6
III.II	CONCEPTO DE REGRESIÓN.....	7
III.III	INFERENCIA BAYESIANA.....	8
III.IV	REGRESIÓN DEL PROCESO GAUSSIANO .....	10
III.V	FUNCIONES DE COVARIANZA (KERNELS) .....	11
III.VI	REGRESIÓN DEL GP A PRIORI .....	16
III.VII	PREDICCIÓN DE LOS PROCESOS GAUSSIANOS.....	18
III.VIII	HIPERPARAMETRIZACIÓN .....	20
IV	CAPITULO II – VIDA Y NO-VIDA.....	23
IV.I	TARIFICACIÓN ANUAL RENOVABLE .....	23
IV.II	TRIANGULO DE PROVISIONES TÉCNICAS .....	25
V	CAPITULO III – APLICABILIDAD EMPÍRICA .....	28
V.I	INTERPOLACIÓN TABLAS DE MORTALIDAD .....	28
V.II	CÁLCULO DE PROVISIONES NO-VIDA .....	32
VI	CAPITULO IV – RESUMEN .....	41
VI.I	CONCLUSIONES .....	41
VI.II	DEBATE Y PRÓXIMOS PASOS .....	41
VI.III	BIBLIOGRAFÍA.....	43
VI.IV	ANEXOS .....	45

# I INTRODUCCIÓN

El actual desarrollo tecnológico de los softwares estadísticos nos permite elaborar nuevas técnicas más complejas, como, por ejemplo, los modelos de Deep Learning o Machine Learning. En este trabajo trataremos uno de estos modelos de aprendizaje supervisado con la finalidad de resolver problemas de regresión, los Procesos Gaussianos (GP) como técnica de predicción.

Los GP son un proceso estocástico no lineal extendidamente conocidos en el ámbito matemático y financiero. Esta técnica aún se encuentra en actual de desarrollo, pues, a día de elaboración de esta actividad, van saliendo nuevas aplicaciones y metodologías de cálculo que afinan el GP, como por ejemplo, en la optimización de hiperparámetros. Su nombre viene dado por Carl Friedrich Gauss, ya que está basado en la noción de una distribución gaussiana o normal, pues en términos generales, un GP es una extensión continua de una distribución gaussiana multivariante.

Tras una exhaustiva explicación del funcionamiento de un GP y cómo funciona la regresión de este, se procederá a la aplicación empírica del GPR a través de dos ejemplos prácticos que pueden encontrarse en el entorno actuarial.

La primera parte de esta aplicación empírica, utilizaremos los GP para obtener la interpolación mensual de las tablas de mortalidad anuales PASEM 2020 Unisex de segundo orden<sup>2</sup>, para poder obtener una mayor precisión en la estimación de la prima en seguros de tarificación anual renovable (TAR), acorde con la edad actuarial.

Por otro lado, aplicaremos la metodología de los GP para la predicción de provisiones técnicas de siniestros pendientes de declaración, realizando una comparativa contra otro método clásico de cálculo de provisiones, el método Chain-Ladder determinista.

De este modo, seremos capaces de modelar funciones no paramétricas que nos permitirán obtener las tareas de predicción e interpolación en términos probabilísticos.

---

<sup>2</sup> PASEM 2020 Unisex de 2º orden se utilizan para los seguros de Vida-Riesgo que no contemplen seguros de decesos.

## II OBJETIVOS Y METODOLOGÍA

Para la realización de esta actividad, se ha utilizado el software RStudio, en su versión 4.1.3 (2022-03-10), un ide de R. Además, también se ha integrado algunas formulaciones extraídas de Stan en nuestro entorno de R, para la inferencia bayesiana<sup>3</sup>.

Esta actividad está compuesta por cuatro capítulos:

En primer lugar, detallaremos los elementos básicos de este proceso estocástico, indagando en los conceptos de como pasamos de unos datos que siguen una distribución normal a un GP, es decir, a obtener una distribución de probabilidades sobre funciones no paramétricas. El siguiente paso en este bloque será una breve explicación de los conceptos previos necesarios para entender en funcionamiento de un GP. Seguido, se realizará un análisis de la necesidad de la obtención de la regresión no lineal, enfatizando en la estadística bayesiana: Teorema de Bayes y la inferencia Bayesiana. Además, procederemos a desarrollar el funcionamiento de la regresión de un GP, con algunas de sus variantes detalladas, profundizaremos sobre las funciones Kernels, pues estas, son las que regirán la calidad del GP, a través de la obtención y optimización de los hiperparámetros, tal y como detallaremos más adelante. Distinguiremos entre las regresiones GP con la integración o no de ruido (media igual o no a cero) y en última instancia, detallaremos en el funcionamiento de la predicción de un GP y la obtención de los hiperparámetros.

El segundo bloque está focalizado en una breve introducción de los conocimientos previos necesarios sobre los dos ejemplos que realizaremos en la aplicación empírica. Por un lado, detallaremos que son los seguros de Vida TAR, que garantía afecta y los posibles beneficios que puede aportarnos la introducción de un GP en este entorno. Por otro lado, también detallaremos que son las provisiones técnicas de los ramos no vida y el desarrollo matemático de algunos métodos de cálculo utilizados para la predicción de estas y sobre que limitaciones se realizarán estos cálculos.

El tercer capítulo recoge la parte empírica de esta actividad. Para el ejemplo de interpolación, se compara el resultado obtenido con el cálculo habitual de la obtención de la prima neta anual. En segundo lugar, con relación a la predicción de provisiones, se realizará una breve explicación del tratamiento inicial que debemos darle a nuestros inputs para integrarlos en la regresión de un GP, siguiendo la misma metodología que en Lally y Hartman, (2018). Se realizará una breve comparativa entre los métodos de obtención de provisiones Chain Ladder determinista contra los obtenidos por las regresiones del GP, siguiendo la misma metodología que Lally y Hartman, (2018) (cálculo de provisiones en base a dos inputs).

En el cuarto y último bloque detallaremos las conclusiones obtenidas de esta actividad, las limitaciones a las que nos hemos ceñido y las próximas líneas de investigación a las que podrían llevar nuestros resultados a un paso más adelante.

---

<sup>3</sup> En el Anexo de este trabajo podrá encontrar el código utilizado para la realización de este.

### III CAPITULO I – PROCESO GAUSSIANO

El GP es una colección de variables aleatorias, tal que, para cualquier subconjunto finito de puntos, estos tienen una distribución conjuntamente gaussiana, donde esta colección finita de puntos sigue una distribución normal multivariada, indexadas por tiempo o espacio. Dicho de otro modo, un GP es una generalización de las distribuciones de probabilidades gaussianas multivariadas.

Previo a indagar en un GP, primero detallaremos brevemente que es una distribución gaussiana y como llegamos a este GP.

#### III.I DISTRIBUCIÓN GAUSSIANA

Una distribución gaussiana es una distribución que queda definida por dos parámetros, la media y su desviación típica, y su función de densidad (la que nos asignará los valores más probables para un conjunto de datos) es la siguiente:

Partiendo de un determinado *input* =  $x$ , la ecuación para una distribución normal univariante tiene la siguiente función de densidad<sup>4</sup>:

Función de densidad de una Distribución Gaussiana Univariante:

$$f(x) = p(y|\mu, \sigma^2) \quad (1)$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

Donde, dominio de  $f = \mathbb{R}$ :

$\mu = E\{x\}$ , siendo este el valor esperado.

$\sigma^2 = Var(x)$ , siendo la varianza.

$\sigma = sd(x)$ , siendo la desviación estándar.

Este tipo de distribución consta de una sola dimensión:  $p = 1$ .  $X \sim N(\mu, \sigma)$ , donde los datos de entrada siguen la distribución de una normal, con dos parámetros, y sean independientes e idénticamente distribuidas (*i.i.d.*).

En el caso del *input* que no sea definido por una sola dimensión, es decir tenga dos o más dimensiones, obtenemos una Distribución Gaussiana Multivariante, siendo una generalización de la distribución normal unidimensional a dos o más dimensiones. Según el Teorema del Límite Central (TLC), la suma de variables aleatorias independientes tiende a una Gaussiana<sup>5</sup>.

$$y \sim N(\mu, \sigma^2) ; \sum_{i=1}^n y_i = N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2) \quad (3)$$

Los parámetros para una distribución multivariante se componen por un vector de medias ( $\mu$ ) y una matriz de varianzas ( $\Sigma$ ), donde se expresa como:  $X \sim N_p(\mu, \Sigma)$ .

<sup>4</sup> Función que nos define, para una variable aleatoria, que valor puede tomar en base a la probabilidad relativa.

<sup>5</sup> Si escalamos una variable gaussiana, también es gaussiana.

Vector  $x = (x_1, \dots, x_p)^T$ , es un vector aleatorio p-dimensional.

Vector de medias:  $E(x) = \mu = (\mu_1, \dots, \mu_p)'$ , donde,  $E(x_i) = \mu_i$

Vector de covarianzas:  $Var(x) = \Sigma$ , cuya posición (i, j) es  $\sigma_{i,j} = Cov(X_i, X_j)$

$Var(x) = E[(X - \mu)(X - \mu)'] = E(XX') - \mu\mu'$

Función de densidad de una Distribución Gaussiana Multivariante:

p-dimensional (2 o más):

$$f_X(x_1, \dots, x_p) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}, x \in \mathbb{R}^p \quad (4)$$

$\mu$  = vector real, siendo la esperanza matemática de X.

$|\Sigma|$  = determinante de la matriz; si es escalar (matriz 1\*1), se reduce a la ecuación de una distribución normal univariante.

La matriz  $\Sigma = AA^T$ ; es la matriz de covarianza de los componentes  $X_j$ . es una matriz real simétrica, semidefinida positiva de dimensión  $p \times p$  (de este modo, sabemos que es invertible). En la diagonal de esta matriz se encuentra las varianzas de las  $X_j$ 's, mientras que alrededor de la diagonal, observamos las covarianzas entre las variables.

$$X \sim N_p(\mu, \Sigma) \rightarrow \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix} \quad (5)$$

Tal y como podemos observar, la definición para un proceso gaussiano y una distribución multivariante es relativamente similar, pero la principal diferencia es que un GP tiene un espacio y un dominio continuo (infinito-decimal), mientras que una distribución multivariante, corresponde con variables aleatorias discretas.

### III.II CONCEPTO DE REGRESIÓN

En términos generales, en una regresión tenemos un conjunto de puntos de datos, donde la finalidad es encontrar una función que nos permita, al introducir los valores de la X (variable exógena), obtener unos datos que sean o más cercanos posibles a Y (variable endógena).

La dificultad de la regresión reside en encontrar una función tal que  $y = f(x)$  dado un conjunto de pares que toman la siguiente forma  $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ .

En este apartado introduciremos una metodología no paramétrica de regresión no lineal, es decir, a diferencia de por ejemplo una regresión lineal, un modelo no paramétrico no tiene una cantidad fija de parámetros definidos. De este modo, podremos encontrar una función que sea capaz de definir una distribución sobre funciones y nos sustentaremos en los intervalos de confianza para la distribución gaussiana, siendo capaces de cuantificar la incertidumbre que pueda existir en la predicción del modelo.

Por ejemplo, la función de la regresión lineal Bayesiana toma la siguiente forma:  $f(x) = w_0 + x^T w$ ,  $y = f(x) + \varepsilon$ , donde supone que la variable  $y$  toma una relación lineal con  $x$ , y presenta ruido independiente con media cero y varianza  $\sigma_n^2$ , donde  $w$  representa los pesos asignados (constante y coeficiente, respectivamente). Si la media es igual a cero, nos quedaría  $f(x) = x^T w$ ,  $\varepsilon \sim N(0, \sigma_n^2)$ .

La finalidad de la regresión del GP es encontrar una función que nos permita encontrar el comportamiento de un conjunto de datos iniciales, y sea lo suficientemente flexible para adaptarse a estos datos. Para lograr este objetivo, un GP utiliza la distribución normal multivariante para encontrar las relaciones de dependencia entre las variables aleatorias, obteniendo un vector de medias y una matriz de covarianzas.

Como método bayesiano, el GP busca las distribuciones posteriores sobre los modelos, las cuales, nos cuantificarán nuestra incertidumbre en los modelos. Los GP entonces se definen como una distribución de probabilidad sobre funciones aleatorias, que nos permite calcular intervalos de incertidumbre. Es decir, partiremos definiendo un GP a priori sobre funciones, y a medida que vayamos introduciendo observaciones, extraeremos la distribución de densidad posterior, manteniendo las propiedades del GP.

Para comprender correctamente esta definición, primero debemos introducir que es una distribución priori y una distribución a posteriori, las cuales se obtienen a través de la inferencia bayesiana.

### III.III INFERENCIA BAYESIANA

En términos generales, la inferencia bayesiana muestra cómo se deben actualizar los datos a medida que se van anidando y observando de nuevos. La inferencia bayesiana parte del Teorema de Bayes, el cual, se basa en conocer la probabilidad condicional de un suceso determinado como A dado B, en el que se analiza la distribución de probabilidad dado A. En forma de ecuación, el teorema queda como:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B \cap A)}{P(A)}$$

La principal ventaja de la inferencia bayesiana es la obtención de las mejores predicciones cuando tenemos una muestra que queremos aprovechar el máximo de esta.

Si aplicamos la regla de Bayes sobre la temática recurrente, nos quedaría la siguiente ecuación:

$$posterior = \frac{verosimilitud * prior}{verosimilitud marginal} \quad p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)} \quad (6)$$

Donde  $w$  es un vector de pesos (parámetros) y  $X$  es el vector input;  $p(w)$  es la distribución a priori;  $p(y|X, w)$  es la verosimilitud;  $p(w|y, X)$  es la distribución a posteriori;  $p(y|X)$  es la densidad marginal de  $X$ :



$$p(y|X) = \int_{\theta} p(y|X, w)p(w)d\theta \quad (7)$$

Observamos como esta ecuación (7) combina los inputs (a priori) y las probabilidades, los dos requerimientos para obtener la distribución posteriori. En los GP, este a posteriori es la media obtenida por la distribución posterior y a la vez, es también su moda, llamado como MAP<sup>6</sup>.

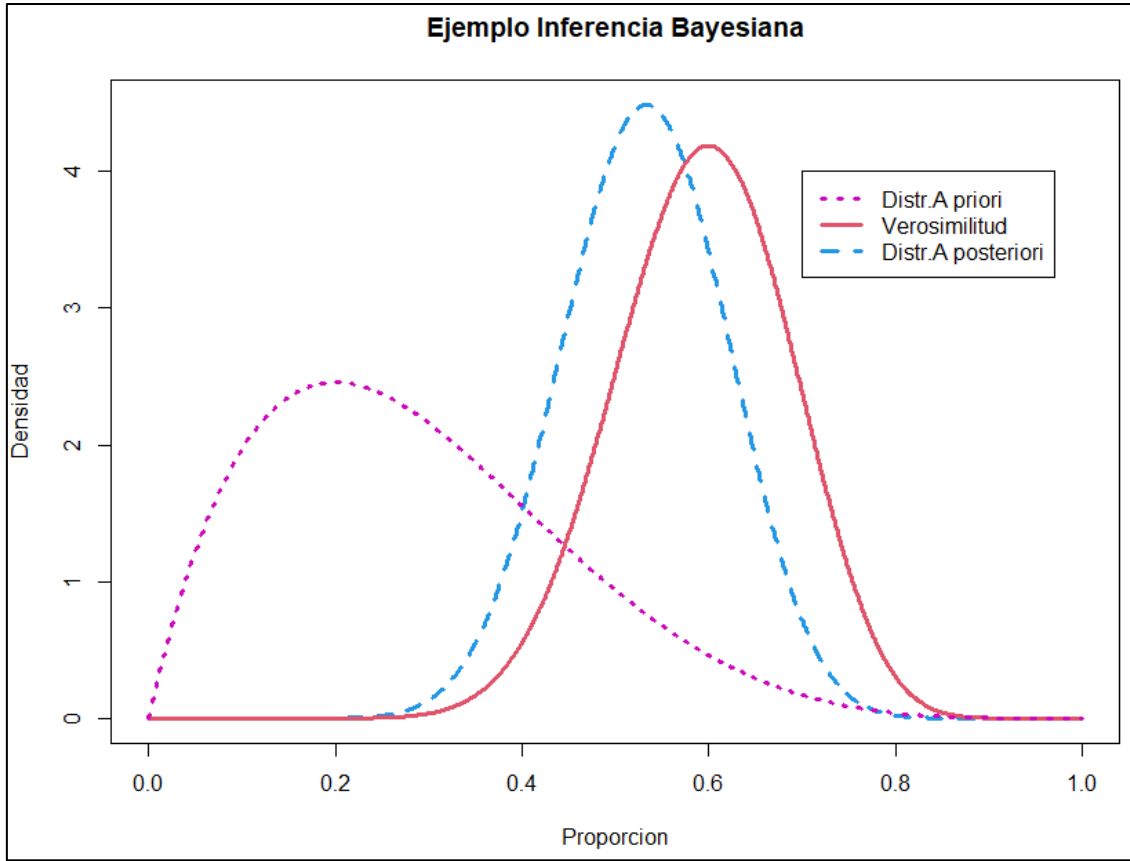
La inferencia bayesiana se basa en tomar las probabilidades como medida de incertidumbre. En este enfoque podemos separar dos tipos de incertidumbre: En primer lugar, la definida como verosimilitud  $p(y|X, w)$ , la cual refleja la incertidumbre aleatoria y no puede reducirse. En segundo lugar, tenemos la distribución a priori  $p(w)$ , la cual, refleja la incertidumbre epistemológica, es decir, esa incertidumbre que podemos ir reduciendo a medida que alimentamos el modelo con observaciones para así obtener  $p(w|y, X)$  a través del Teorema de Bayes.

A modo de ejemplo, supongamos que queremos saber cual es la proporción entre lanzamientos de una moneda, a través de las probabilidades, si son cara o cruz. Definimos previamente que este comportamiento sigue una distribución a priori beta<sup>7</sup>. La función en RStudio para distribuciones continuas es “*dbeta*”, y definimos unos parámetros iniciales de  $\alpha = 2$  y  $\beta = 5$ . Además, sabemos que de 25 lanzamientos, 15 han sido cara y 10 han sido cruz:

---

<sup>6</sup> Estimaciones Máximas a Posteriori (MAP); Más adelante entraremos en detalle sobre esta metodología.

<sup>7</sup> En el capítulo de aplicación empírica se entrará más en detalle sobre la función Beta y sus aplicaciones.



*Figura 1: Elaboración propia con RStudio. Inferencia bayesiana para la obtención de una distribución a posteriori adaptada a los datos de entrada.*

Tal y como podemos observar, inicialmente partimos de una distribución beta, y, una vez extraída la verosimilitud observamos que se distancia en gran proporción de la distribución definida inicialmente, por este motivo, podemos observar, después de aplicar la inferencia bayesiana, que la distribución a posteriori ya no se asemeja a una distribución beta, sino, más bien, a una distribución normal o gaussiana. A partir de esta función a posteriori que se asemeja a una normal, podríamos proceder a extraer las predicciones para los próximos lanzamientos de la moneda en base a las probabilidades.

### III.IV REGRESIÓN DEL PROCESO GAUSSIANO

En un GP se asume que un proceso subyacente  $f(x)$  sigue una distribución gaussiana, donde esta distribución viene definido por los parámetros de la media y covarianza.

Definidos  $m(x)$  como la función de media y  $k(x, x')$  como la función de covarianza. Un ejemplo matemático viene dado por Rasmussen et al., (2005), el cual, toma la siguiente forma:

$$m(x) = E[f(x)], \quad (8)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] = Cov(x, x'), \quad (9)$$

$$f(x) \sim GP(m(x), k(x, x')) \quad (10)$$

Donde, para un conjunto finito de datos de entrada (inputs), la función marginal toma la siguiente forma:

$$f(x) \sim N(m(x), k(x, x')) \quad (11)$$

La función obtenida es la distribución a priori, la cual, a medida que vayamos integrando los inputs, la regresión se adaptará a la forma y comportamiento de estos, obteniendo la distribución a posteriori.

Cuando la media toma valor igual a cero, es decir, la media de las variables aleatorias son cero, la función viene explicada enteramente por la función Kernel seleccionada a priori, de forma que cualquier función Kernel puede usarse como función de covarianza. Una función Kernel, en términos estadísticos son funciones predeterminadas que se asignan para definir la matriz de covarianzas. A continuación, entraremos en detalle sobre estas funciones Kernel, pudiendo observar que tan relevantes son para la regresión de un GP.

### III.V FUNCIONES DE COVARIANZA (KERNELS)

Tal y como hemos comentado previamente, primero definiremos un GP a priori sobre funciones. El GP a priori vendrá definida por la selección del Kernel que definiremos a continuación. Comúnmente, se define que la función de la media  $m(x)=0$ , de este modo, solo deberíamos, por el momento, definir una función de covarianza o Kernel. Más adelante observaremos como es la resolución introduciendo una media distinta a cero.

En términos estadísticos, un Kernel o núcleo es un método no paramétrico que nos permite estimar la función de densidad de probabilidad de una variable aleatoria, a partir de un número finito de observaciones. La estimación de densidad que obtenemos es una medida relativa de verosimilitud, de este modo, cuanto mayor sea su valor, mayor evidencia de que la observación pertenece a la distribución.

A diferencia de un histograma convencional, que agrupa por intervalos, el Kernel nos permite obtener las distribuciones continuas, una para cada observación de la muestra. La agregación de las contribuciones individuales entre el total de observaciones permite obtener una curva final que describe la distribución conjunta de la muestra. Es importante, a la hora de estimar un Kernel, la selección del ancho de banda ( $\ell$ ), es decir, que tan suavizada debe ser cada una de las distribuciones marginales. El ancho de banda ( $\ell$ ) es un parámetro que se encuentra en todas las funciones de Kernel (excepto el Kernel Polinomial) y definirá la forma de este. Hay varios criterios para encontrar el valor óptimo de este parámetro, tal y como detallaremos más adelante.

Este ancho de banda controla cuanto se puede expandir la influencia de cada observación. Cuando se selecciona un ancho de banda grande, se necesitarán de un elevado número de observaciones para poder calcular la estimación en cada punto, lo que puede hacer aumentar nuestro sesgo de estimación<sup>8</sup> al definir excesiva suavidad. En el caso contrario,

---

<sup>8</sup> El sesgo de estimación es la diferencia entre el valor esperado del estimador contra el valor real de este parámetro.

si se selecciona un ancho de banda estrecho, las observaciones necesarias son inferiores, lo que puede comportar en un aumento de la variabilidad y falta de suavidad.

Cuando se emplea un Kernel con distribución normal, este valor  $\ell$  es igual a la desviación típica, y los pesos asignados siguen la distribución de una normal estándar:  $K(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$

Es importante destacar que la función de covarianza (Kernel), define la forma y la importancia de los pesos que se asocian a cada observación para el cálculo de la estimación en un GP.

Hay algunos métodos para estimar el mejor tipo de Kernel y ancho de banda que se ajusten a nuestros datos, los comúnmente más usados son, en primer lugar, la validación cruzada<sup>9</sup>, aunque requiere de un mayor coste computacional, y en segundo lugar, por la máxima verosimilitud. En este trabajo, analizaremos la mejor estimación través de la máxima verosimilitud bayesiana.

Por último, la función de densidad para un conjunto de variables aleatorias independientes e idénticamente distribuidas, su estimador de densidad de Kernel es:  $\hat{f}_\ell(x) = \frac{1}{n\ell} \sum_{i=1}^n K\left(\frac{x-x_i}{\ell}\right)$ , donde K es el núcleo.

Tal y como hemos observado hasta ahora, la regresión del GP nos permite obtener funciones que nos definan la función general de una regresión, pero esta última, se verá modificada en base a la selección del Kernel que utilizamos para la obtención de esta, y más, cuando no hay ruido blanco, es decir, la media es igual a cero, ya que la función del input dependerá en su totalidad de la función asignada al Kernel.

### III.V.I TIPOS DE KERNEL

Definido  $x = \{x_1, \dots, x_p\}$  y el correspondiente output  $f(x)$ :  $cov(f(x), f(x')) = k(x, x')$ , siendo una matriz simétrica definida positiva.

A continuación, se ejemplifican algunos de los kernels más comúnmente usados:

Kernel Exponencial Cuadrado (SE) o (RBF):

$$k_{SE}(x, x') = \sigma^2 \exp\left(-\frac{1}{2\ell^2} \|x - x'\|^2\right) \quad (12)$$

Donde  $\sigma^2$  es la varianza (diagonal de la matriz<sup>10</sup>).

Kernels Mattern [Matérn, 1960]

Los kernels de clase Matérn vienen dados por:

<sup>9</sup> La validación cruzada consiste en una técnica que evalúa los resultados de análisis estadísticos que garanticen independencia entre los datos de entrenamiento (training) y los datos de prueba (test), en base a varios escenarios planteados (múltiples iteraciones).

<sup>10</sup> Normalmente se emplea con un valor de  $\sigma^2 = 1$ , de este modo, el valor es constante.

$$k_{\text{Matérn}}(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2} \|x - x'\|}{\ell} \right)^\nu K(\nu) \left( -\frac{\sqrt{2} \|x - x'\|}{\ell} \right) \quad (13)$$

Los más comunes en ámbitos de Machine Learning son los siguientes:

$$k_{\text{Matérn}}^{1/2}(x, x') = \exp \left( -\frac{\|x - x'\|}{\ell} \right) \quad (14)$$

$$k_{\text{Matérn}}^{3/2}(x, x') = \left( 1 + \frac{\sqrt{3} \|x - x'\|}{\ell} \right) \exp \left( -\frac{\sqrt{3} \|x - x'\|}{\ell} \right) \quad (15)$$

$$k_{\text{Matérn}}^{5/2}(x, x') = \left( 1 + \frac{\sqrt{5} \|x - x'\|}{\ell} + \frac{5 \|x - x'\|^2}{3\ell^2} \right) \exp \left( -\frac{\sqrt{5} \|x - x'\|}{\ell} \right) \quad (16)$$

Rational Quadratic (RQ):

$$k_{\text{RQ}}(x, x') = \sigma^2 \left( \frac{1 + \|x - x'\|^2}{2\alpha\ell^2} \right)^{-\alpha} \quad (17)$$

Donde  $\alpha$  es un parámetro de variación de escala. Cuando  $\alpha$  tiende a infinito, el Kernel RQ se asemeja a un Kernel SE a medida que alpha toma valores crecientes.

Periodic Kernel (PK):

$$k_{\text{PK}}(x, x') = \sigma^2 \exp \left( -\frac{2\sin^2(\pi \|x - x'\|/p)}{\ell^2} \right) \quad (18)$$

Donde  $p$  es un parámetro que controla el período<sup>11</sup> de la función.

Linear/Polynomial Kernel (LPK)

$$k_{\text{LPK}}(x, x') = (x^T x' + \sigma^2)^d \quad (19)$$

Donde, para el conjunto de Kernels,  $\|x - x'\| = d$ , siendo  $d$  la distancia euclidiana entre dos valores.

Estos kernels nos dan como salida una matriz de covarianzas  $K$  de tamaño  $[n, n]$ , la cual, nos describe la covarianza entre dos inputs.

Sin añadir ninguna observación, los resultados serían nuestras distribuciones bayesianas a priori, basadas exclusivamente en el diseño del modelo. Es una suposición de cómo se distribuyen conjuntamente los valores de la función subyacente que tratamos de modelar. Sabiendo que  $X_1$  y  $X_2$  son fijos, podemos decir que el GP a priori nos describe una

---

<sup>11</sup> Podría interpretarse de forma similar cuando definimos la estacionariedad en un modelo ARMA[p].

distribución sobre funciones, el cual, es una distribución gaussiana multivariada de  $p$  dimensiones.

A medida que se van añadiendo las observaciones de la muestra, iríamos obteniendo las distribuciones a posteriori condicionadas a estas observaciones.

Los resultados inferenciales de estas ecuaciones dependerán de los valores de los denominados hiperparámetros, por ejemplo,  $\alpha$  y  $\ell$ , ya comentados, los cuales, definirán el comportamiento del modelo. Uno de los métodos más utilizados en el ámbito estadístico para la obtención del valor óptimo de estos hiperparámetros es a través de las estimaciones máximas a posteriori (MAP) con algunas selecciones a priori. Es decir, la estimación óptima de los hiperparámetros puede obtenerse a través del enfoque de máxima verosimilitud.

Por último, cabe decir que la obtención de la matriz de covarianzas no se limita a estos kernels, pues también existe la posibilidad de realizar operaciones con kernels, es decir, se pueden combinar diferentes funciones de covarianza en base a la obtención de un mejor modelo. Por ejemplo:

La suma de las funciones de covarianza es una función de covarianza válida:

$$k_{new}(x, x') = \sum_i k_i(x, x') \quad (20)$$

De la misma forma, el producto de las funciones de covarianza es una función de covarianza válida:

$$k_{new}(x, x') = \prod_i k_i(x, x') \quad (21)$$

A continuación, visualizaremos el impacto de la definición de estos parámetros sobre cada una de las funciones Kernel que hemos detallado previamente. Para esta visualización de distribuciones a prior y sin introducción de ruido, definimos un vector de valores  $X$  comprendido entre -1 y 1, de longitud 200 observaciones, usando la función “*mvrnorm*”<sup>12</sup> crearemos cinco distribuciones que siguen una normal multivariante, modificando los parámetros respectivos de cada función:

---

<sup>12</sup> En el capítulo de aplicabilidad empírica entraremos en detalle sobre los argumentos de esta función.

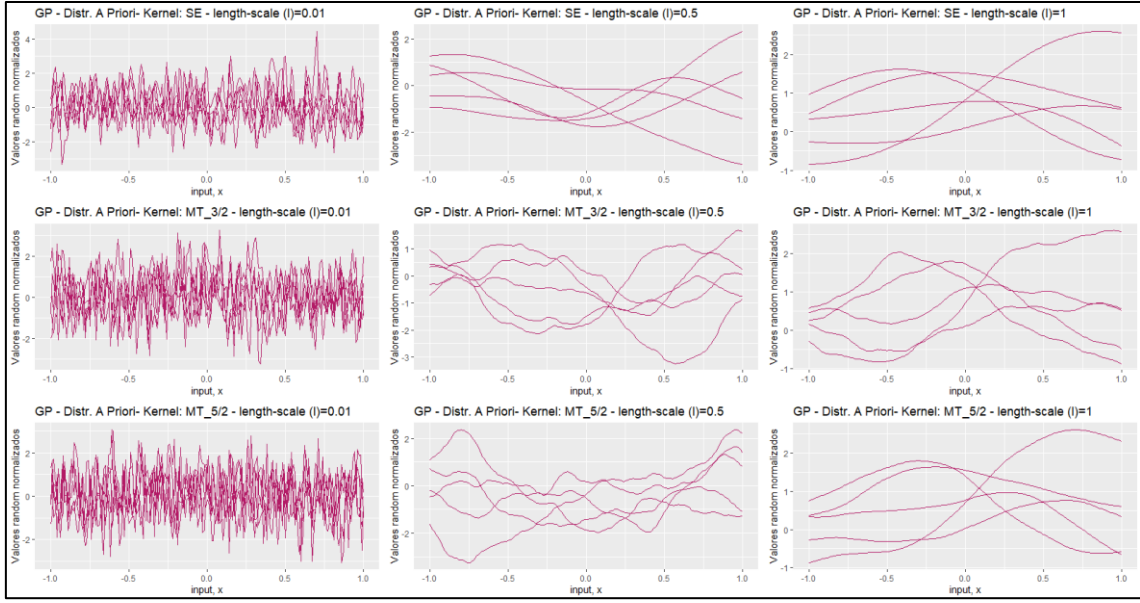


Figura 2: Elaboración propia con RStudio. Distribuciones a priori de los siguientes kernels en base a la modificación del parámetro length-scale: Exponencial, Mattern 3/2 y Mattern 5/2.

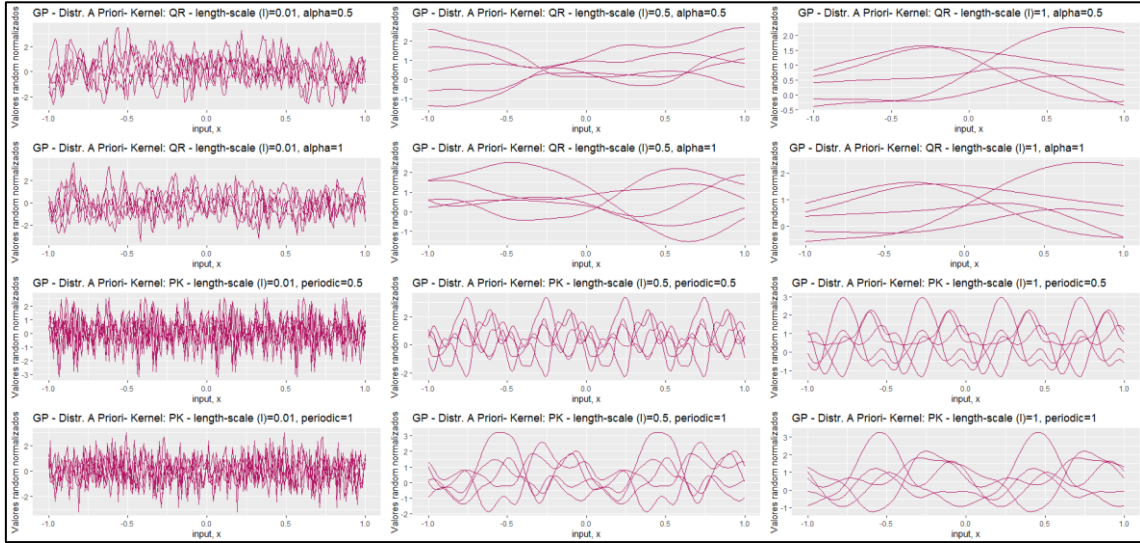


Figura 3: Elaboración propia con RStudio. Distribuciones a priori de los siguientes kernels en base a la modificación del parámetro length-scale, alpha y periodicidad: Racional Cuadrático y Periódico.

Tal y como podemos observar en las figuras 2 y 3, tanto la elección del Kernel como el valores de los parámetros de este influirán significativamente en la función de distribución. A medida que aumentamos el valor de  $\ell$ , observamos un comportamiento común en todos los Kernels, y es, tal y como hemos comentado previamente, un suavizado de nuestras distribuciones a priori. También podemos apreciar como las funciones Mattern recogen una mayor irregularidad en su continuidad (curvas) respecto el resto de las funciones. Por otro lado, podemos observar claramente en la figura 3 el funcionamiento del Kernel periódico y como repite su comportamiento y como se distancian las curvas periódicas al modificar el parámetros periódico. Algo similar ocurre

con el Kernel racional cuadrático, donde al modificar el parámetro alpha, y definir un ancho de banda reducido, toman mayores diferencias, en cambio, para un ancho de banda igual a 1, el cambio de valor del parámetro alpha afecta en menor medida, el cual, a medida que aumenta su valor, aumenta la similitud con el Kernel Exponencial.

Esta visualización gráfica previa nos permite, en un inicio y a criterio del especialista, tratar de seleccionar esos Kernels que mejor se puedan adaptar a los datos de entrada, y, en base a la dimensión y tipología de estos, la selección de los valores de los parámetros óptima. De todos modos, en el apartado de hiperparametrización, detallaremos otras metodologías para obtener el valor del parámetro óptimo adecuado a los inputs.

### III.VI REGRESIÓN DEL GP A PRIORI

Una vez tengamos definido el Kernel que vamos a utilizar para definir nuestra matriz de covarianzas, el siguiente paso es ir añadiendo la información, es decir, integrar las observaciones.

A continuación, montamos la ecuación que nos permitirá observar cómo se estructura cada una de las funciones que debemos definir como continuas, siempre y cuando las funciones de media y auto covarianzas sean continuas también. Podemos observar en la siguiente ecuación  $[X_1, f(X_1)]$  como el par de entradas y  $X_2$  como el conjunto que queremos realizar la predicción a  $N$  puntos.

$$\begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} \sim N \left( \begin{bmatrix} m(X_1) \\ m(X_2) \end{bmatrix}, \begin{pmatrix} k(X_1, X_1) & k(X_1, X_2) \\ k(X_2, X_1) & k(X_2, X_2) \end{pmatrix} \right) \quad (22)$$

$X_1$  es una matriz de tamaño  $N \times P$ , siendo  $N$  el recuento de datos de entrenamiento y  $P$  la dimensión donde se ubican estos datos. Por otro lado,  $X_2$  es una matriz de tamaño  $M \times P$ , siendo  $M$  el número de ejemplos a predecir.

#### III.VI.I PROCESOS GAUSSIANOS CON MEDIA IGUAL A CERO (SIN RUIDO)

Por simplicidad de cálculo definimos  $m(X_1)$  y  $m(X_2)$  iguales a cero, la distribución a priori de la función de densidad gaussiana queda como:

$$p \left( \begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} \right) = \frac{1}{(2\pi)^{\frac{n_1+n_2}{2}} \det(k)^{1/2}} \exp \left( -\frac{1}{2} \begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} K^{-1} \begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} \right) \quad (23)$$

Siendo  $K^{-1}$  la inversa de la matriz.

Por otro lado, la distribución a posteriori del proceso gaussiano  $f(X_2)$  es:

Ecuación a Posteriori:

$$f(X_2) | f(X_1), X_1, X_2 \sim N (k(X_2, X_1)k(X_1, X_1)^{-1}f(X_1)) \quad (24)$$



$$k(X_2, X_2) - k(X_2, X_1)k(X_1, X_1)^{-1} k(X_1, X_2)) \quad (25)$$

### III.VI.II PROCESOS GAUSSIANOS CON MEDIA DISTINTA A CERO (CON RUIDO)

Hasta ahora hemos tratado el GP siendo la media cero, aunque esta suposición no siempre se cumple, con lo cual, realizaremos un breve desarrollo para esas situaciones donde la media realmente no es cero.

La integración de ruido en un GP no trata de modificar la ecuación, sino que, partiendo de la ecuación resultante para un modelo sin ruido, se añadirán unos parámetros de suponer que la media no es igual a cero.

Si al igual que una regresión lineal,  $y = f(x) + \varepsilon$ , añadimos a nuestra ecuación del GP ruido independiente, con varianza  $\sigma_n^2$ , nos quedaría la siguiente ecuación:

$$f(x) \sim GP(m(x), k(x, x')) + \sigma_n^2 \delta_{xx'} \quad (26)$$

Donde  $\delta_{xx'}$  es un valor delta de Kronecker que toma valor cero si  $x \neq x'$ , y valor uno en caso de que  $x = x'$ . Nótese  $m(x) \neq 0$ , para valores de  $x$ .

Ecuación a Priori:

$$\text{cov}(y_1, y_2) = k(X_1, X_2) + \sigma_n^2 \delta_{xx'}, \text{ o igualmente, } \text{cov}(y) = k(X_1, X_2) + \sigma_n^2 I_N \quad (27)$$

$$\begin{bmatrix} f(X_1) \\ f(X_2) \end{bmatrix} \sim N \left( \begin{bmatrix} m(X_1) \\ m(X_2) \end{bmatrix}, \begin{pmatrix} k(X_1, X_1) + \sigma_n^2 I_N & k(X_1, X_2) \\ k(X_2, X_1) & k(X_2, X_2) \end{pmatrix} \right) \quad (28)$$

Ecuación a Posteriori:

$$f(X_2) | f(X_1), X_1, X_2 \sim N (k(X_2, X_1)(k(X_1, X_1) + \sigma_n^2 I_N)^{-1} f(X_1) \quad (29)$$

$$k(X_2, X_2) - k(X_2, X_1)(k(X_1, X_1) + \sigma_n^2 I_N)^{-1} k(X_1, X_2)) \quad (30)$$

Observamos que al introducir solo los términos adicionales de suposición de existencia de ruido, las matrices de covarianza solo se modifican al añadir estos parámetros, pero no la estructura en sí. Por ejemplo, para el Kernel exponencial cuadrático quedaría de la siguiente forma:

$$k_{SE}(x, x') = \exp \left( \frac{-1}{2\ell^2} \|x - x'\|^2 \right) + \sigma_n^2 \delta_{xx'} \quad (31)$$

Una vez definida la ecuación posterior, ya podemos obtener la media posterior y la covarianza.

### III.VII PREDICCIÓN DE LOS PROCESOS GAUSSIANOS

La predicción en los GP se basa en dos partes. En primer lugar, el valor esperado de la media y en segundo lugar en base a la matriz de covarianza asignada (Kernel).

En primer lugar, la media posterior es un vector de longitud  $n$ , mientras que la matriz de covarianzas es una matriz de tamaño  $n \times n$ , donde, dado el GP a priori y una verosimilitud gaussiana, la función posterior también es un GP. A continuación, podemos observar la ecuación de la obtención de las funciones de predicción cuando los hiperparámetros son fijos.

$$m_{2|1} = m(X_2) + k(X_2, X_1)(k(X_1, X_1) + \sigma_n^2 I_N)^{-1}(f(X_1) - m(X_1)) \quad (32)$$

$$\Sigma_{2|1} = k(X_2, X_2) - k(X_2, X_1)(k(X_1, X_1) + \sigma_n^2 I_N)^{-1}k(X_1, X_2)^T \quad (33)$$

Estas expresiones que acabamos de definir no tienen aplicación hasta que previamente hayamos encontrado, en un inicio, el valor óptimo de los hiperparámetros, pero dado un conjunto de observaciones y dada una función de covarianza previamente asignada, ya podríamos obtener la densidad posterior.

Nótese como la estimación de la media posterior es una combinación lineal de las observaciones de la función  $f(x)$ , mientras que la varianza es independiente de las observaciones de la función  $f(x)$ .

Tener en cuenta que en el caso que la estimación sea  $m(x)=0$ , tan solo haría falta extraer de las ecuaciones el término  $\sigma_n^2 I_N$ .

Tomando en cuenta las dos últimas ecuaciones definidas previamente, la distribución de predicción final, con integración de ruido, quedaría como:

$$Y(x_*) \sim N(m_*(x_*), \Sigma_{2|1}(x_*) + \sigma^2(x_*)) \quad (34)$$

Donde  $Y(x_*)$  podría definirse como nuestra función final.

Posteriormente, entraremos en el proceso de entrenamiento de nuestro GP, es decir, el procedimiento para la obtención de los hiperparámetros (parámetros del Kernel y la varianza del ruido si precede).

A continuación, realizaremos un pequeño ejercicio donde puede visualizarse la obtención de la distribución a posteriori a medida que vamos introduciendo nuestros inputs en la regresión del GP. Para este ejemplo usaremos un Kernel Exponencial, aplicado a los datos utilizados para la obtención de las figuras 2 y 3. Ahora, en vez de crear 5 distribuciones, crearemos 100, para obtener un intervalo de confianza más ajustado entre punto y punto y poder recoger una mayor variabilidad de la varianza. Como hemos visto hasta ahora, nuestra función a prior dependía enteramente solo de nuestros valores de  $X$ , el siguiente

paso será ir introduciendo en nuestra regresión los valores de  $Y$  observados. Por ejemplo, queremos predecir los valores comprendidos entre  $X = -1$  y  $1$ , y nuestros valores observados (*train*) son los siguientes:  $x_i = (-0.81, -0.58, -0.2, 0.33, 0.51, 0.9)$ , por otro lado, nuestros valores conocidos de  $Y$  son los siguientes:  $y_i = (3.21, 1.8, 0.2, -1.9, 1.75, 2.1)$ . Por último, hemos definido un ancho de banda de  $0.1$  y un valor para la integración de ruido de  $0.2$ .

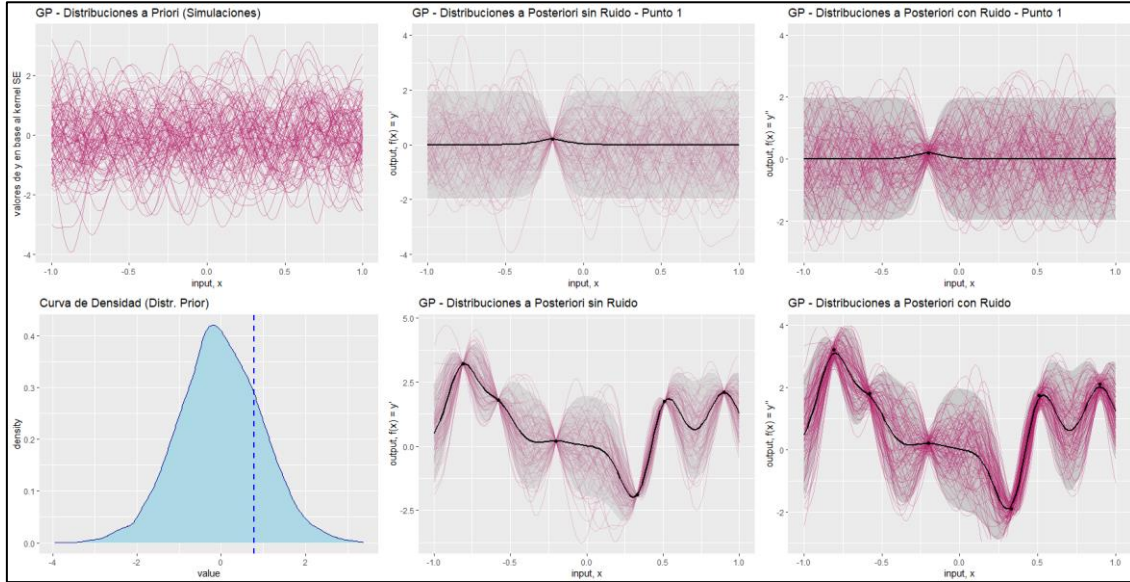


Figura 4: Elaboración propia con RStudio. Visualización de una regresión por proceso gaussiano, con Kernel exponencial, ancho de banda definido igual a  $0.1$  y ruido  $0.2$ . El intervalo de confianza asignado tiene un nivel de confianza asignado del 95% (bandas grises).

En la figura 4 podemos observar, en primer lugar, la distribución a priori, armada a través de la realización de cien distribuciones que siguen una normal multivariante que depende completamente del Kernel exponencial. Las dos últimas figuras de la primera fila se visualizan como se adapta la regresión a medida que integramos puntos observados de nuestro input, en este caso, perteneciente al  $x_3 = -0.2$  y a  $y_3 = 0.2$ , con y sin ruido en la regresión. A medida que vayamos integrando los puntos, nuestra distribución a posteriori se irá adaptando paulatinamente a las variables observadas, definiendo también los intervalos de confianza. La primera figura de la segunda fila es un histograma de nuestra distribución a priori, pudiendo observar el punto medio y cuáles son esos valores que más se repiten, reflejándose que, con los parámetros y variables definidas, nuestra regresión tiende a cero. La segunda figura de la fila dos es la regresión del GP sin ruido con toda la muestra introducida, por otro lado, la última figura es la regresión del GP con ruido. La primera diferencia destacable que podemos observar es que, cuando la media no es cero, la regresión no estrangula tanto los puntos (efecto cuello de botella), y se ensancha el margen de actuación de la media<sup>13</sup>. Por otro lado, podemos observar que, como la media tiende a cero en los tramos con pocas observaciones, la regresión puede tomar valores con una tendencia que no sea correcta entre los puntos de la muestra. Por este motivo, es importante realizar un proceso de entrenamiento del modelo. Una de las metodologías para evitar que suceda esta casuística es definir una función para nuestros

<sup>13</sup> Podemos observar que la integración con ruido permite un mayor margen de maniobra en el valor esperado de la media, siendo un método indicado para esas regresiones que requieren una mayor flexibilidad tanto en observaciones conocidas como para predecir.

parámetros, evitando que tomen valores constantes a lo largo del período de muestra.<sup>14</sup> Otra posible solución es con la integración de ruido, que permitirá al modelo una mayor flexibilidad de adaptación. Cabe decir que este comportamiento vendrá condicionado a la distancia entre los puntos de la muestra, es decir, si se observa una alta varianza entre dos puntos muy cercanos, cuando previamente no se había observado este comportamiento a lo largo del camino, en los siguientes puntos de la muestra el modelo estima con esta alta volatilidad, forzando la regresión a fuertes caídas y subidas.

### III.VIII HIPERPARAMETRIZACIÓN

La correcta elección de los valores de los parámetros nos permitirá explicar mejor los datos de entrenamiento. Para la obtención de estos hiperparámetros óptimos entramos en un proceso de aprendizaje de parámetros, en el cual, se irán modificando los valores de estos parámetros hasta alcanzar un modelo que se adapte lo suficientemente bien para explicar los datos de entrenamiento.

Algunos de los hiperparámetros observados hasta ahora son  $\sigma_n^2$  o  $\ell$ , los cuales agruparemos como  $\theta$  para facilitar la metodología explicativa de estos hiperparámetros.

La obtención de esta optimización puede obtenerse por varios criterios en los GP, el método que trataremos en esta sección es por selección bayesiana.

#### III.VIII.I VEROSIMILITUD DE LA SELECCIÓN BAYESIANA

En términos generales, el criterio de máxima verosimilitud (MV) toma como objetivo encontrar el modelo que, con una mayor probabilidad, haya generado en la máxima similitud, los datos observados. Es un modelo generativo para series de tiempo.

Para nuestro caso, un modelo generativo es cualquiera que sea capaz de modelar la variable output como una variable aleatoria  $Y$  a través de una distribución condicionada a los inputs de entrada  $X$  y a los parámetros  $\theta$ .

Tal y como hemos observado previamente, la regla genérica de bayes para realizar inferencia bayesiana es:

$$posterior = \frac{verosimilitud * prior}{verosimilitud marginal} \quad (35)$$

Como la obtención de la verosimilitud marginal puede implicar una integral de difícil resolución, la metodología tiene a derivarse por aproximación de máxima verosimilitud. Para los GP de media cero, la verosimilitud de los parámetros se monta de la siguiente forma.

En primer lugar, si observamos la probabilidad marginal  $p(y|X, \theta, \sigma)$ :

$$p(y|X, \theta, \sigma) = \int p(Y|f, X, \theta, \sigma)p(f|X, \theta, \sigma)df \quad (36)$$

<sup>14</sup> En el capítulo de aplicabilidad empírica, observaremos como llevar a cabo este proceso.

$$p(y|X, \theta, \sigma) = \frac{1}{(2\pi|K_y|)^{n/2}} \exp\left(-\frac{1}{2}(y - m)^T K_y^{-1}(y - m)\right) \quad (37)$$

Donde  $m = m(X)$  y  $K_y = K_\theta(X, X) + \sigma_n^2 I_N$

Esta ecuación nos permitirá adaptar a una Gaussiana, aunque sea continua, no nos afectará a la hora de la resolución, ya que los datos que disponemos son variables discretas que siguen una distribución multivariante. Cada distribución marginal de cada predicción obtenida (*output*) es una gaussiana univariante.

Como la finalidad de esta resolución es la obtención de un modelo con probabilidad máxima, de forma que maximizaremos la probabilidad marginal con respecto a los parámetros del modelo. En este caso, como la probabilidad marginal es una función exponencial, podemos tomar logaritmos para contrarrestar la exponencial y aprovecharemos que la función logarítmica aumenta estrictamente, con lo cual, el output obtenido por  $\log(p(y|X))$  son iguales a  $p(y|X)$ , obteniendo el mismo valor para los parámetros en los dos casos.

A continuación, extraemos el Negative Log-Likelihood (NLL), de este modo, quedará definida nuestra función objetivo al minimizar la log-verosimilitud que equivale a maximizar la exponencial.

$$NLL = -\log P(Y|X, \theta) \quad (38)$$

$$NLL = -\frac{1}{2} \log(|K_y|) - \frac{1}{2} (y - m)^T K_y^{-1} (y - m) + \frac{n}{2} \log(2\pi) \quad (39)$$

En orden lineal, la primera parte de la ecuación (39) nos definirá en qué sentido está orientada la función, la cual, penaliza la complejidad, es decir, es un término de penalización por el error de un modelo al ajustar los valores observados que depende solamente del Kernel y las entradas. Por otro lado, la parte intermedia de la ecuación nos indica la distancia orientada de la media, es similar a la distancia euclidiana para contemplar el ajuste de los datos. Es otro término de penalización que aumentará proporcionalmente ante la complejidad de un modelo y es la única que depende de las observaciones de  $Y$ . En último lugar, es un término constante de normalización, que no modifica el valor del resultado. De este modo, a través de esta ecuación, ya podríamos tratar de encontrar cual es la función que tiene un mayor comportamiento similar a la media de todos los puntos. Podremos observar que a mayor ancho de banda ( $\ell$ ), menos complejas son las funciones, son más suaves, con lo cuál, también se vuelven más rígidas, provocando que no se ajuste correctamente a los datos si estos son muy dispersos. En resumen, la ecuación da un mayor peso a las funciones que se ajusten correctamente a los datos sin ser tan complejas.

Para el desarrollo empírico de nuestra regresión según el GP, seguiremos la misma metodología que (Flaxman et al., 2015) es decir, la inferencia bayesiana, el cual, nos permite definir una distribución conjunta de probabilidad  $p(\cdot)$  para los hiperparámetros de la función de covarianza:

$$\theta \sim p(\theta) \quad (40)$$

$$f|X, \theta \sim N_n(M(X), K_\theta(X, X)) \quad (41)$$

Donde  $K_\theta(X, X)$  es una función del input de nuestros datos y de los hiperparámetros. En vez de tomar un valor fijo o constante, los hiperparámetros de la función de covarianza vendrán definidos por una distribución, de la familia gaussiana, por ejemplo: exponencial, gamma, log-normal, etc.<sup>15</sup>

A continuación, realizaremos una breve introducción sobre, en primer lugar, los seguros de tarificación anual renovable (TAR) en Vida, para el escenario que nos es de interés, y en segundo lugar, de la estructura y algunas técnicas para el cálculo de provisiones técnicas (PT) en seguro No-vida.

---

<sup>15</sup> En el capítulo de aplicación empírica entraremos en detalle sobre las funciones de parámetros.

## IV CAPITULO II – VIDA Y NO-VIDA

### IV.I TARIFICACIÓN ANUAL RENOVABLE

Un seguro<sup>16</sup> de tipo TAR en vida-riesgo<sup>17</sup> son los seguros que ofrecen cobertura para una sola anualidad y que pueden modificarse a fecha de renovación. Al ser anuales, la prima también tiende a ser de pago anual, la cual queda sujeta a incrementarse en base a la edad del asegurado, a la siniestralidad histórica, a la revalorización de los capitales contratados y a otros índices económicos como el tipo de interés máximo o la actualización a nuevas tablas de mortalidad designados por la Dirección General de Seguros y Fondos de Pensiones (DGSFP) y el Boletín Oficial de Estado (BOE), respectivamente. Hay otros factores internos que pueden modificar la cuantía de dicha prima como los gastos internos o de administración (GGI) y los gastos externos o comerciales (GGE), así como otros factores que componen las bases técnicas (BT) de una póliza que queden sujetos a modificarse.

Para nuestra actividad, nos ceñiremos al siguiente escenario: La obtención de las tasas de mortalidad se extraerá de las tablas de mortalidad PASEM2020\_General\_2ndo\_orden<sup>18</sup>, Unisex, destinada a los seguros de vida-riesgo, excluyendo cualquier cobertura por seguro de decesos. El ejercicio solo contemplará la garantía de fallecimiento por cualquier causa, el cual, incluye cobertura para todo tipo de fallecimiento, aunque se excluyen algunos sucesos como las casuísticas que ya recoge el Consorcio de Compensación de Seguros o el fallecimiento bajo efectos de embriaguez. Cabe decir que cada compañía difiere en algunos de los artículos del condicionado general, y adaptará las coberturas en base al perfil del asegurado en las condiciones particulares y/o especiales.

El cálculo que nos permitirá obtener las tasas de mortalidad y así la prima neta, parte de los valores de la cohorte  $l_x$ . Estos valores son reales positivos  $\iota: [0, \omega] \rightarrow \mathbb{R}$ , y representan el monto de personas supervivientes nacidos en el mismo instante, desde el nacimiento hasta la extinción de este grupo de personas Ayuso et al., (2007).  $\iota(0)$  es el tamaño inicial, donde las tablas que usaremos equivalen a 1.000.000 personas y cada  $l_x$  es el número de supervivientes en cada tramo de edad igual a  $x$ . Por otro lado, el cálculo de probabilidad de fallecimiento,  $q_x$ , para este escenario, donde  $h=1$ , siendo  $h$  el período de interés, el tanto anual de mortalidad se calcula a través de la siguiente ecuación:<sup>19</sup>

$$q_x = \frac{(l_{x-1} - l_x)}{l_{x-1}} \quad (42)$$

Esta ecuación equivale a la variación del valor de cohorte por tramo de edad. El ejercicio consistirá en obtener la interpolación mensual de cada  $q_x$  anual, y donde el pago, en caso de que el suceso se de a cabo, es a mes vencido.

Cabe comentar que, bajo la perspectiva de una compañía aseguradora, solo nos focalizaremos en el intervalo de edad comprendido entre 15 y 70 años, simulando la

---

<sup>16</sup> Entorno de cobertura ofrecida en España.

<sup>17</sup> Incluyen esos seguros que dan cobertura para vida y accidentes.

<sup>18</sup> [https://www.boe.es/diario\\_boe/txt.php?id=BOE-A-2020-17154](https://www.boe.es/diario_boe/txt.php?id=BOE-A-2020-17154)

<sup>19</sup> Para una mayor extensión de información, consultar Ayuso et al., (2007).

situación de que el producto viene ofrecido de una aseguradora con destinación hacía una compañía o colectivo laboral, y no a particulares.

Una vez obtenidas las probabilidades de fallecimiento, a continuación, construiremos las ecuaciones para obtener la prima neta por asegurado. Definidas unas bases técnicas y un tipo de interés ( $I_1$ ) anual, y para un intervalo de interpolación definido cada 12 períodos<sup>20</sup>:

$$Tasas puras = q_x * (1 + I_1)^{1/12} \quad (43)$$

Donde  $(1 + i)^{1/12}$  representa la actualización financiera mensual.

$$Tasas Tarifa = \frac{Tasas puras}{(1 - GGI - GGE)} \quad (44)$$

Donde GGI y GGE son los gastos internos y externos expresados en términos porcentuales sobre la prima.

$$Prima Neta = Capital * Tasas Tarifa \quad (45)$$

El capital es el monto acumulado que deberá abonar el asegurador al asegurado en caso de que este último fallezca.

Estas expresiones previas, podemos agruparlas en la siguiente ecuación:

$$\frac{m/V^{(h)}A_{\overline{x}:\overline{n}}}{[u(t - mh)]} = \sum_{t=mh}^{(m+n)h-1} u(t - mh)(1 + I_1)^{-(t+1/12)/h} \left( \frac{l_{xh+t}^{(h)} - l_{xh+t+1}^{(h)}}{l_{xh+t}^{(h)}} \right) \quad (46)$$

Donde  $u(t - mh)$  es el capital esperado asegurado, siendo  $t$  el período, en este caso es igual a 1,  $m$  el diferimiento, donde para nuestro ejercicio no hay, es igual a cero,  $V^{(h)}A_{\overline{x}:\overline{n}}$  el valor actuarial de las primas puras, (en esta ecuación se aplican los recargos sobre la prima pura en vez de sobre la tasa pura (el resultado es el mismo)),  $n$  es el período de finalización de la cobertura, igual a 1 en este caso,  $h$  es la frecuentación de pago, en este caso es igual a 1, e  $I_1$  es el tipo de interés anual, el cual se actualizará mensualmente. Esto último ya es una ventaja teniendo en cuenta que, normalmente, el tipo de interés se actualiza o a mitad de la anualidad o al finalizar la anualidad, en base a cuando se realice el pago, en el caso que se dé el fallecimiento.<sup>21</sup>

Como podemos observar, el cálculo es sencillo y no se han incluido cobertura o factores adicionales como un recargo estándar, participación por beneficios, descuentos sobre garantías, grupos de riesgo o recargos por mayores coberturas como la enfermedad profesional, ya que tan solo nos interesa el comportamiento de  $q_x$  y no una simulación real en sí. Por este motivo, nos focalizaremos sobre la prima neta, sin entrar en detalle sobre el cálculo de la prima total, una vez se han integrado los impuestos necesarios como el del consorcio o el I.P.S.

<sup>20</sup> Interpolación para 12 períodos dentro de cada anualidad, obteniendo así la actualización financiera mensual. Se ha realizado la interpolación a nivel mensual por criterio y por limitación de coste computacional.

<sup>21</sup> En términos continuos, se tiende a tomar logaritmos sobre el tipo de interés y se multiplica por la ecuación con variables discretas:  $\frac{i_h}{Ln(1+i_1)} * m/V^{(h)}A_{\overline{x}:\overline{n}}$ .



## IV.II TRIANGULO DE PROVISIONES TÉCNICAS

Las provisiones técnicas (PT) son el importe que deben mantener las entidades aseguradoras y de reaseguro que garantizan todas las obligaciones derivadas de los contratos de seguros y que se reflejan en el balance. Estas PT's evidencian la estabilidad de la entidad frente a oscilaciones aleatorias o cíclicas de siniestralidad y posibles riesgos especiales.

En este reporte solo introduciremos la metodología de cálculo de las PT para los seguros de No-Vida. No entraremos en detalle sobre la normativa estipulada el Solvencia II para simplificar los cálculos, ya que la finalidad de la actividad es observar cómo se adapta una regresión GP (con varios kernels) contra los métodos convencionales de cálculo de PT, utilizando algunos de los métodos que si requieren esta normativa.

El método seleccionado es Chain-Ladder determinista, a través de la predicción de triángulos de provisiones.

Limitaciones: No se contemplará el cálculo de las PT con la integración de los tipos de interés ni de la inflación. Tampoco tendremos en cuenta otros elementos que la legislación de Solvencia II contempla, como importes mínimos, cálculos simplificados u opciones transitorias, por ejemplo.

Las PT, según la normativa de Solvencia, están formadas por la suma de la mejor estimación (BEL: el valor actual esperado de los flujos de caja futuros) y el margen de riesgo (margen que garantiza que el valor de las PT sea igual al importe que las entidades previamente exigirían para poder asumir y garantizar las obligaciones de los contratos). Pueden calcularse en su conjunto o por separado (defecto), en esta actividad, solo nos focalizaremos en el cálculo por separado de los siniestros pendientes de declarar en términos brutos.

En este trabajo, analizaremos las reservas por pérdidas como un triángulo de provisiones, para el caso de los siniestros IBNR (*Incurred But Not yet Reported*). Los siniestros IBNR son esos siniestros que ya han ocurrido pero la reclamación aún no se ha tramitado y su comunicación y pago se realizará en un escenario futuro, en varios períodos, con la finalidad de compensar las pérdidas sufridas con pagos por pérdidas en varios momentos. El triángulo de provisiones está compuesto por el importe agregado total de la cartera de una aseguradora de los siniestros IBNR.

El triángulo<sup>22</sup> se estructura por el eje horizontal, llamado año de desarrollo, el cual, equivale al retraso de los pagos de todos los siniestros de la cartera. Por otro lado, el eje vertical pertenece al año de ocurrencia de los siniestros.

---

<sup>22</sup> Estos triángulos de provisiones también reciben la nomenclatura de triángulos *run-off* o de desarrollo.

Ejemplo de un triángulo de PT para siniestros pendientes de declaración en términos brutos de un producto No-Vida, no acumulados:

Año Origen	Año Desarrollo									
	$j_1$	$j_2$	$j_3$	$j_4$	$j_5$	$j_6$	$j_7$	$j_8$	$j_9$	$j_{10}$
$i_1$	$c_{1,1}$	$c_{1,2}$	...	...	...	...	...	...	...	$c_{1,10}$
$i_2$	$c_{2,1}$	$\ddots$								
$i_3$	$\vdots$		$\ddots$							
$i_4$	$\vdots$			$\ddots$						
$i_5$	$\vdots$				$\ddots$					
$i_6$	$\vdots$					$\ddots$				
$i_7$	$\vdots$						$\ddots$			
$i_8$	$\vdots$							$\ddots$		
$i_9$	$\vdots$								$\ddots$	
$i_{10}$	$c_{10,1}$									$c_{10,10}$

*Triangulo 1: Elaboración propia. Plantilla de un cuadro de provisiones técnicas.*

Donde:

Eje horizontal (j): Año desarrollo respecto siniestros ocurridos en año origen.

Eje vertical (i): Año origen del siniestro.

$\mathcal{D} = \{i, j | i \in I, j \in J\}$ , dominio con esquema triangular.

La Diagonal (**negrita**) representa: La parte izquierda, los datos reales actuales de los que disponemos; La parte derecha de la diagonal, los valores a predecir.

$c_{ij}$ : cuantía abonada en el año de desarrollo (j) respecto a los siniestros ocurridos en el año de origen (i).

$C_{ij}$ : cuantía acumulada abonada en el año de desarrollo (j) respecto a los siniestros ocurridos en el año de origen (i).<sup>23</sup>

#### IV.II.I CHAIN LADDER DETERMINISTA

La metodología Chain-Ladder (CL) es uno de los métodos más comúnmente usados en el ambiente actuarial para el cálculo de PT no Vida debido a su sencillez y eficacia para la predicción.

Este método se estructura de la siguiente forma:

En primer lugar, sobre un triángulo de provisiones con valores acumulados por filas ( $C_{i,j}$ ), se extraen los factores de desarrollo a través de la siguiente formulación:

$$F_j = \frac{\sum_{i=1}^{k+1-j} C_{i,j}}{\sum_{i=1}^{k+1-j} C_{i,j-1}} ; \forall j = 2, \dots, k. \quad (47)$$

<sup>23</sup> Sumatorio acumulado por filas.

Donde  $F_j$  será cada factor de desarrollo que usaremos a continuación. Estos factores nos describen cual será la evolución en su conjunto de los pagos por cada par de columnas. En términos generales, la proyección se obtendrá del producto de estos factores por el valor previo por fila acumulada:

$$\tilde{C}_{i,k-i+1+h} = C_{i,k-i+1} \prod_{m=1}^h F_{k-i+1+m} \quad (48)$$

Una vez extraídas las nuevas cuantías acumuladas (última columna del triángulo), tan solo haría falta desacumular de nuevo los valores del triángulo por filas (tomar diferencias) y extraer el sumatorio por diagonales, obteniendo el valor total de provisiones a reservar para poder hacer frente las pérdidas esperadas.

## V CAPITULO III – APLICABILIDAD EMPÍRICA

### V.I INTERPOLACIÓN TABLAS DE MORTALIDAD

Con la finalidad de obtener nuestras nuevas tasas de mortalidad, nuestro input se comprende de dos variables: Por un lado, tenemos la edad anual ( $x$ ) y por otro lado los valores de la cohorte<sup>24</sup> a cada tramo  $l_x = y$ , el cual, deberemos estandarizar para integrarlo correctamente en la regresión del GP, tal y como se observa en la ecuación (49).

$$l'_i = \frac{l_i - \bar{l}}{sd(l)} \quad (49)$$

La edad anual la mantendremos en su formato original ya que no nos repercute en este ejercicio. Para la elaboración de esta parte de la actividad hemos creado una tabla con los datos de cinco asegurados con datos aleatorios que contiene tres variables: El identificador del asegurado, el monto de capital asegurado y la fecha de nacimiento de cada uno de estos. En este escenario, hemos definido las siguientes BT's: GGI=10%, GGE=10%, la fecha de efecto es el 01/05/2022 y  $I_1 = 0.46\%$ <sup>25</sup>

El siguiente paso es extraer la edad actuarial de cada asegurado, calculado de la diferencia entre la fecha de efecto común y la fecha de nacimiento individual y dividido entre 365.25, de este modo, recogemos el impacto de los años bisiestos. Normalmente, una vez obtenida la edad actuarial se redondean los decimales a la unidad en base a si el decimal es inferior o superior a  $x,5$ . Para nuestro ejemplo, mantendremos los dos decimales, ya que será el indicativo para saber a qué mensualidad pertenecen.

Procedemos a iniciar la interpolación por la regresión de un GP de nuestros datos. Previo, cabe destacar los Kernels que hemos utilizado para esta parte de la actividad: Kernel Exponencial, Mattern 3/2, Mattern 5/2, Racional Cuadrático. Se han descartado el Kernel periódico y el Kernel polinomial ya que, debido a la tipología de nuestros datos, estos no se ajustan.

Para la obtención de nuestras simulaciones se ha usado la función “*mvrnorm*”<sup>26</sup> del paquete “MASS”. Esta función nos permite crear una muestra de datos con distribución normal multivariante con parámetros  $\mu$  y  $\sigma$  definidos, definiendo 50 caminos aleatorios que siguen una normal multivariante.<sup>27</sup> A continuación, visualizaremos las distribuciones a prior en base al tipo de Kernel seleccionado, la definición de los valores de los hiperparámetros y de las distribuciones en base si integramos o no ruido.

Cabe comentar que para este apartado de la actividad, no se ha estimado el valor óptimo de los parámetros, pues se han seleccionado a criterio del investigador en base a la

<sup>24</sup> En el Anexo de este trabajo se encuentra el detalle con los valores de cohorte y una tabla adicional con el detalle de los datos de los asegurados aleatorios.

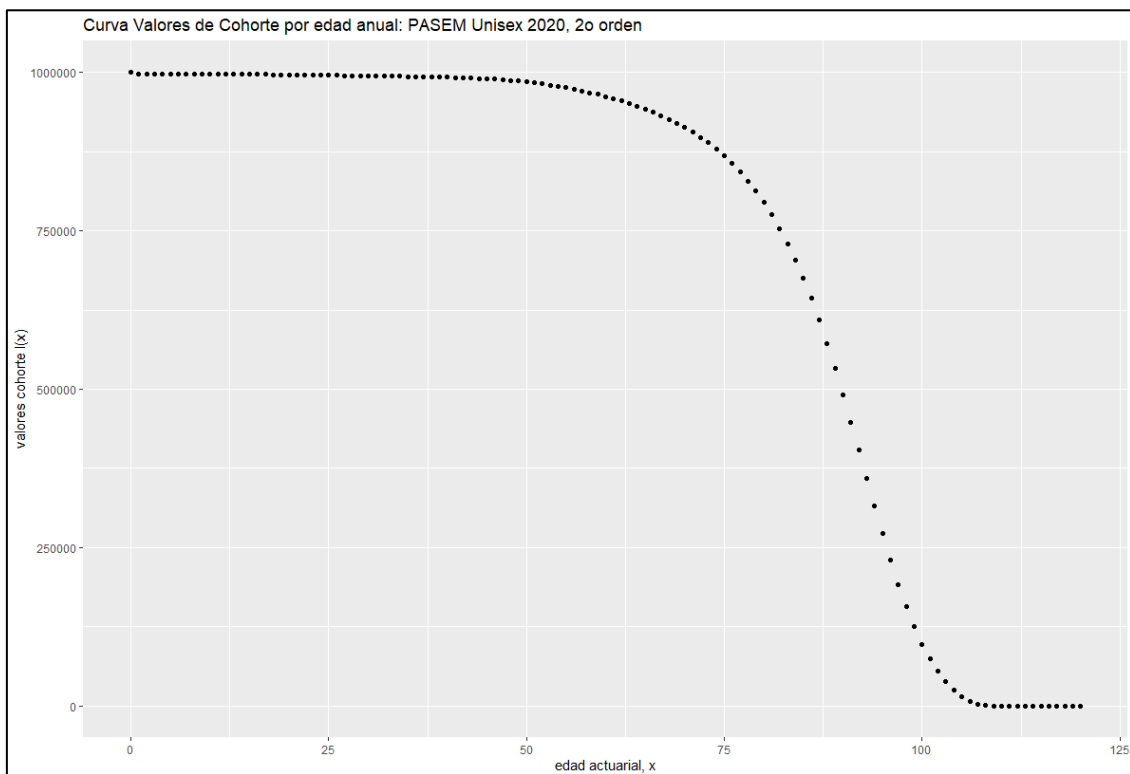
<sup>25</sup> <http://www.dgsfp.mineco.es/es/Regulacion/DocumentosRegulacion/Resolucion%20tipo%20de%20intereses%20de%20seguros%202022.pdf>

<sup>26</sup> Esta función realiza la descomposición de la matriz de covarianzas a través de los valores propios (*eigen values*), es más estable que la descomposición de Choleski, pero también más lenta computacionalmente.

<sup>27</sup> Parte del código utilizado para esta aplicación empírica se ha basado en el blog online: (Bottiger, 2012).

tipología de los datos de entrada y a un análisis de prueba-error. El parámetro ancho de banda seleccionado es un valor constante igual a 1, al igual que el parámetro alpha, mientras que por otro lado, el parámetro de ruido se ha definido con un valor de 0.15, también constante.

Nuestros inputs X son una secuencia de valores (edad) que van desde 0 a 120 años, los valores Y son los valores de cohorte para cada tramo de edad y la predicción será los valores de Y' que se encuentran a nivel mensual, es decir, X' es la interpolación de X mensual ( $X' = X * 1/12$ ).



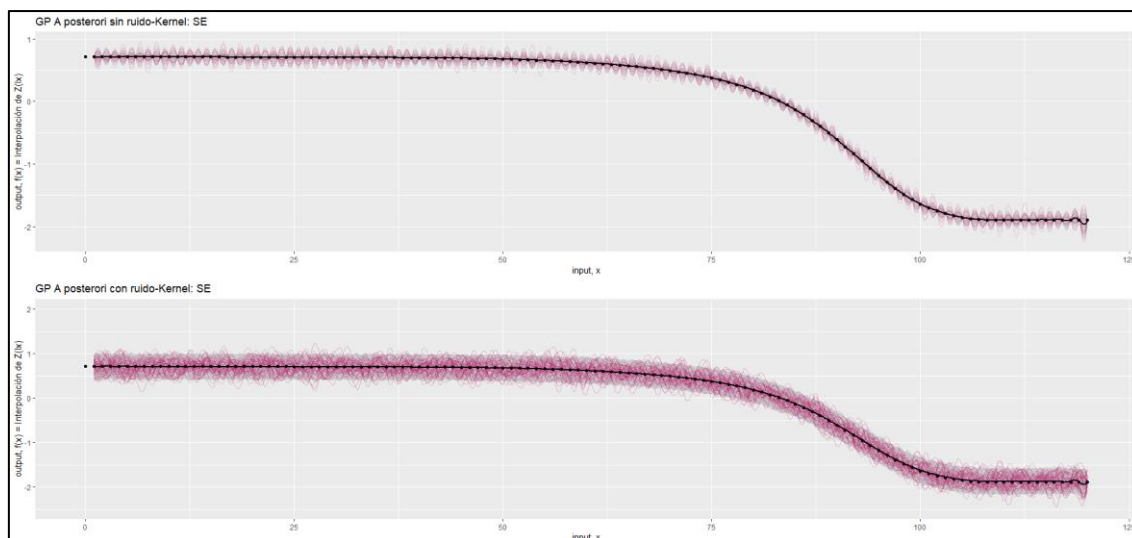
*Figura 5: Elaboración propia con RStudio. Valor de cohorte comprendido entre la edad de 0 años a 120 años, para una muestra de 1.000.000 personas, perteneciente a la tablas PASEM Unisex 2020 de 2no orden.*

Previo a realizar la predicción, previamente, si visualizamos los valores de cohorte graficados en la figura 5 podemos sacar las siguientes conclusiones: Observamos que la reducción del monto de población decrece muy levemente hasta, aproximadamente 50 años. A partir de los 50, cae progresivamente hasta los 100 aproximadamente, llegando a 0 supervivientes alrededor de los 110 años. Con estos datos, podemos extraer ya la conclusión que las probabilidades de fallecimiento crecerán exponencialmente a partir de los 50 años, penalizando el valor de las primas notablemente.

Tras realizar un primer análisis para la selección del Kernel óptimo, hemos podido extraer las siguientes conclusiones:

En primer lugar, seleccionamos el Kernel Exponencial como distribución a priori en base a la tipología de los datos. Se descartan los Kernel tipo Mattern debido a que presentan demasiada volatilidad entre cada punto (edad actuarial), y nos interesa obtener una media constante que enlace cada punto dividiendo entre 12 subperíodos. Por último, se descarta

el Kernel racional cuadrático ya que por la tipología de los datos, este presenta, la curva cuadrática entre punto y punto, y si queremos suavizar este comportamiento debemos aumentar el valor de  $\alpha$  hasta alcanzar una curva continua, pero al realizar este proceso, cada vez tomara una mayor similitud con el Kernel exponencial.<sup>28</sup>



*Figura 6: Elaboración propia con RStudio. Interpolación del Valor de cohorte comprendido entre la edad de 0 años a 120 años, para una muestra de 1.000.000, perteneciente a la tablas PASEM Unisex 2020 de 2no orden, con Kernel Exponencial y parámetro ancho de banda constante igual a 1 y valor sigma igual a 0.15.*

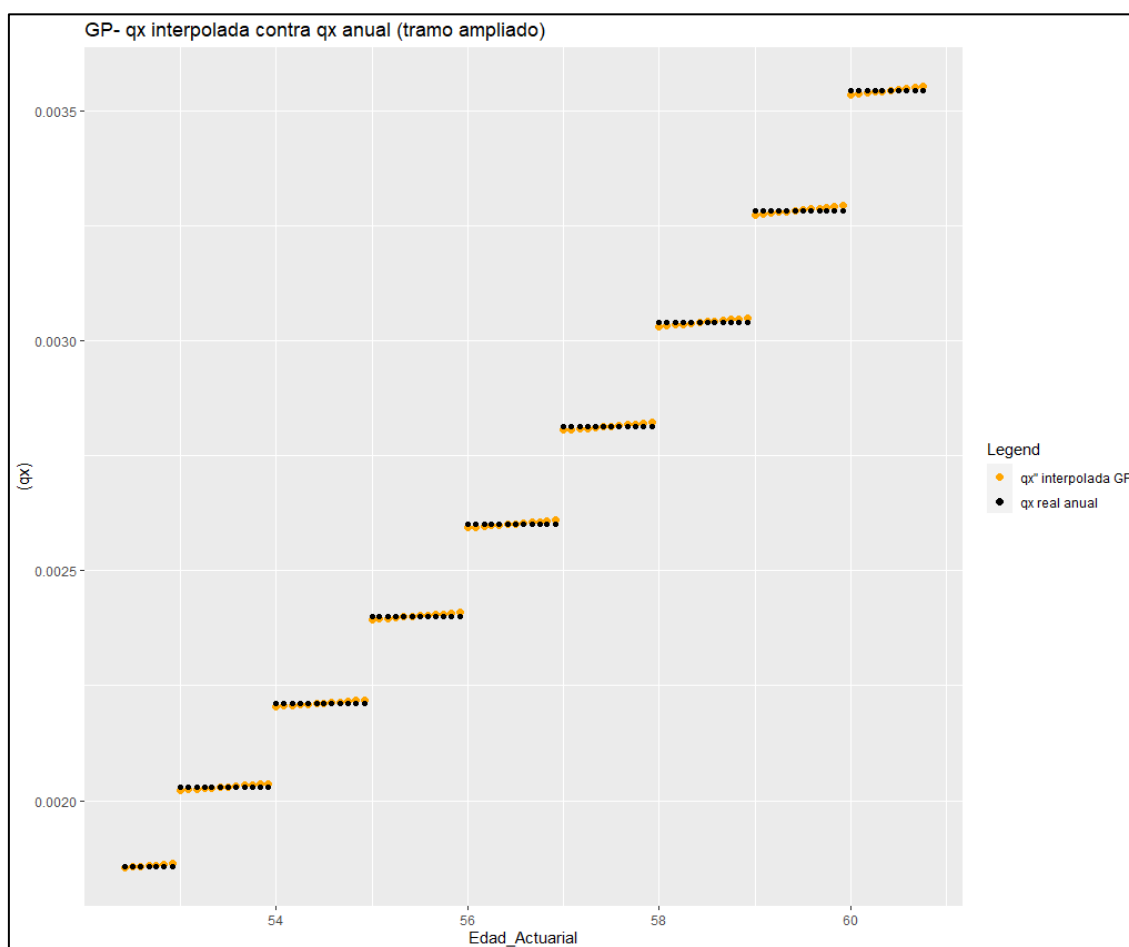
En la figura 6 podemos observar la distribución a posteriori, con y sin la integración de ruido del Kernel exponencial. En primer lugar podemos visualizar que en el último tramo, toma pequeñas desviaciones (edad 118 a 120) respecto a la continuidad de estos puntos, esto se debe a que como a partir de 112 años, el valor de cohorte es igual a cero hasta 120 años, el modelo estima de todos modos una varianza superior a cero, forzando estas desviaciones. Con la finalidad de arreglar esta pequeña desviación, habría que revisar el modelo y entrenarlo en una mayor profundidad, pero como solo nos centraremos en los tramos de edad comprendidos entre 15 y 70 años, no nos afecta. En segundo lugar, la integración de ruido en este modelo, podemos observar que no es de interés para nuestro salida de datos deseada, pues no requerimos de flexibilidad en la media, sino que recoja los puntos exactos, es decir, que tenga la mayor continuidad lineal entre cada punto, por este motivo, realizaremos el cálculo de la prima con el Kernel exponencial sin la integración de ruido.

En siguiente paso es desestandarizar los valores de cohorte interpolados por GP y extraer las probabilidades de fallecimiento ( $q_x$ ) para poder calcular la prima neta en base al desarrollo ya comentado en el capítulo anterior. Una vez obtenidas las nuevas  $q_x$ , se ha propuesto el siguiente tratamiento. El cálculo habitual de la obtención de las tasas de mortalidad corresponde a la variación de los valores de cohorte (ecuación 42), si esta variación es ahora mensual, obtendremos las primas mensuales, pero como la finalidad de esta actividad es calcular la variación de la prima anual, se propone el siguiente cálculo para obtener la prima anual en base a la edad actuarial con corte mensual:

<sup>28</sup> Por criterio de sostenibilidad, no se adjunta el detalle gráfico, pero puede encontrarse en el script de este trabajo.

$$q''_x = q'_x + q_x * \left(1 - \frac{1}{12}\right) \quad (50)$$

Siendo,  $q''_x$  la probabilidad de fallecimiento que aplicaremos sobre cada asegurado;  $q_x$  es la probabilidad de fallecimiento anual (calculada sobre los valores de cohorte originales);  $q'_x$  es la probabilidad de fallecimiento obtenida por interpolación cada doce períodos. De este modo, recogemos el peso que tiene la probabilidad de interpolación sobre la probabilidad real. Dicho de otro modo, podemos introducir correctamente el cálculo para un seguro de vida, con cobertura anual (prima se abona a inicio del contrato) pero con pago vencido mensual (abono de la aseguradora al asegurador del capital asegurado en caso de fallecimiento). A continuación realizaremos un análisis gráfico para entender mejor este último concepto y observar los beneficios de esta interpolación:



*Figura 7: Elaboración propia con RStudio. Comparativa entre las probabilidades de fallecimiento anuales contra las probabilidades de fallecimiento interpoladas y tratadas para el tramo de 54 a 60 años.*

En la figura 7 observamos la diferencia entre como afecta, la probabilidad de fallecimiento cuando el pago es anual vencido, con actualización mensual, que tomaría un valor constante para todo el año, contra la probabilidad de fallecimiento con pago mensual vencido y actualización mensual, donde observamos que la probabilidad es creciente, encontrándose por debajo de la  $q_x$  hasta la mitad del período aproximadamente, y un valor superior al final de este, ajustando correctamente y más apropiadamente en

base a la edad actuarial del asegurado. A continuación visualizaremos una tabla con los datos de los asegurados para observar esta diferencia en términos de prima neta:

<i>id_aseg</i>	<i>Nacimiento</i>	<i>Edad Act</i>	<i>Edad "</i>	<i>Edad</i>	<i>Capital</i>	<i>PN" (INT)</i>	<i>PN (Real)</i>	<i>DIF PRIMAS</i>
<i>aseg_1</i>	07/09/1981	40,65	40,67	41	25.000	11,33	12,46	-1,13
<i>aseg_2</i>	01/04/1959	63,08	63,08	63	100.000	557,56	558,83	-1,27
<i>aseg_3</i>	30/01/1971	51,25	51,25	51	50.000	105,48	105,64	-0,16
<i>aseg_4</i>	15/07/1992	29,79	29,83	30	35.000	7,70	7,61	0,09
<i>aseg_5</i>	14/12/1969	52,38	52,42	52	150.000	347,99	348,08	-0,09
<b>Total =</b>						1.030,06	1.032,61	-2,55

*Tabla 1: Elaboración propia. Comparativa entre las primas netas calculadas por la metodología tradicional contra la nueva metodología estimada por GP. Edad y PN son las columnas calculadas por la metodología tradicional y Edad'' y PN''(INT) por la nueva metodología.*

En la tabla 1 disponemos de varios contenidos: la Edad Act es la edad actuarial calculada según la fecha de efecto tomando en cuenta el año bisiesto. Por otro lado, Edad'' es el tramo de edad por interpolación al que se le aplicará la tasa, mientras que Edad es el tramo de edad por anualidad al que se le aplicará la tasa a cada asegurado, respectivamente. Podemos observar que ahora podemos encontrar unas tasas más personalizadas respecto la edad actuarial, donde se puede reflejar en la columna Edad''. Realmente, si observamos la columna PN(real) y PN''(INT), podemos ver que no hay una gran diferencia sobre la prima total en el acumulado final, pero si cabe decir que esta metodología se adapta levemente mejor a las necesidades del asegurado en el período de cobertura. Observamos que en los tramos de edad que se encuentran por debajo de la mitad decimal de su edad actuarial, la nueva prima neta es más barata respecto la prima total común. Por otro lado, podemos comentar que tampoco es un patrón común para todos los tramos de edad, pues por ejemplo, en el primer asegurado, podemos observar que el asegurado, al tener una edad actuarial de 40.65, en el cálculo convencional se le aplicarían las tasas pertenecientes a la edad de 41 años, mientras que con la nueva metodología, le pertenece el tramo de 40.67 años, el cual, supone un ahorro para el asegurado.

## V.II CÁLCULO DE PROVISIONES NO-VIDA

Nuestro input consta de dos variables exógenas y una variable endógena. Las variables exógenas pertenecen a los años de inicio del importe de los siniestros y el año de desarrollo de estos, mientras que la variable endógena pertenece al cúmulo del importe de estos propios siniestros. De este modo, nos quedaría un plano espacial de tres dimensiones, donde las dos variables exógenas serían la latitud y la longitud y las pérdidas los valores en cada momento del tiempo: Tanto los años de origen como de desarrollo son una matriz de tamaño de 10x10, dando un total de 100 observaciones. 55 observaciones de muestra y 45 observaciones a predecir.

### V.II.I TRATAMIENTO INPUT (*warping input beta*)

Para nuestra actividad, tal y como hemos observado hasta ahora, deberemos darle tratamiento a nuestros inputs con la finalidad que tengan una media de cero y una



desviación estándar igual a la unidad. Seguiremos la metodología empleada en Lally y Hartman, (2018), donde el año de origen y el año de desarrollo serán tratados de la siguiente forma.

Tras observar nuestros datos, podemos observar una falta de estacionariedad, con lo cual, en línea con la transformación aplicada en Snoek et al., (2014) modificamos nuestros datos de entrada con la finalidad de obtener una mayor flexibilidad en la definición de creencias previas sobre formas particulares de no estacionariedad. De este modo, el input debe ser normalizado deformado a través de una función de distribución acumulativa beta Snoek et al., (2014), obteniendo una función que si presenta estacionariedad, modificando la función de covarianza de  $k(x_p, x_q)$  a  $k(w(x_p), w(x_q))$ : Lally y Hartman, (2018) y Ruitenberg, (2019).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (51)$$

$$\omega_j(x_{ij}^*) = \int_0^{x_{ij}^*} \frac{\mu^{\alpha_j-1} (1-\mu)^{\beta_j-1}}{B(\alpha_j, \beta_j)} d\mu \quad (52)$$

Donde  $x_{ij}^*$  es igual a  $x'$ , los inputs normalizados de los años de origen y de desarrollo, y cada  $\omega_j$  es una función de distribución acumulada (CDF) comprendido entre cero y uno.

El principal beneficio de una función beta es que puede tomar diferentes formas fácilmente solo modificando los parámetros de alpha y beta, y con el input normalizado, se puede aplicar al encontrarnos en un baremo entre cero y uno, dándonos la flexibilidad necesaria a la hora de adaptar la función.

Por otro lado, los valores (volumen de pérdidas acumuladas en nuestro caso) de la variable endógena, se estandarizarán (normalizan) con la siguiente formulación:

$$Z'_i = \frac{Z_i - \bar{Z}}{sd(Z)} \quad (53)$$

Siendo  $\bar{Z}$  el promedio del vector (media) y  $sd(Z)$  la desviación estándar de la muestra. Como nuestro input estará normalizado, Lally y Hartman, (2018) propone la integración del parámetro Señal del GP ( $\eta^2$ ). Este parámetro multiplica cada uno de los Kernels con la finalidad de evitar que la diagonal de la matriz, cuando la distancia euclídea (input de nuestro modelo) entre  $x_p$  y  $x_q$  sea igual a cero, la diagonal de la matriz tomará valor 1 y con la integración de este parámetro, da una mayor flexibilidad al modelo.

Al igual que con la metodología de Chain-Ladder, la predicción se obtendrá a partir de las pérdidas acumuladas por año de desarrollo.

El input (X) contiene dos variables, año desarrollo y año origen, que puede definirse como un plano de dos dimensiones, donde:

$D = (\omega_{i1}, \omega_{i2}, z_i)_{i=1}^n$ ,  $\omega_1$  pertenece al año de origen y  $\omega_2$  al año de desarrollo.

Así pues, nos quedaría la siguiente forma para nuestro input:

$$\omega_1(x) = \text{BetaCDF}(x, \alpha_1, \beta_1) \quad (54)$$

$$\omega_2(x) = \text{BetaCDF}(x, \alpha_2, \beta_2) \quad (55)$$

El siguiente y último paso, previo a la revisión de los datos de salida, es la definición de la distribución que van a seguir en un inicio todos los parámetros a estimar, una vez ya hemos definido a su vez, la distribución que seguirán nuestros datos input.

## V.II.II DISTRIBUCIONES PRIOR E HYPERPARAMETROS

Una vez tenemos los inputs tratados, el siguiente paso sería escoger/seleccionar cual es la distribución que deben seguir estos inputs. Lally y Hartman, (2018) y Ruitenberg, (2019) aplican algunas modelizaciones como la distribución de t\_Student (con 4 grados de libertad, media cero y varianza 1), gamma (4,4; parametrizado con forma y ratio con valor 4) y log-normal (con media cero y desviación estándar 0.5). Estas distribuciones están supeditadas a ser igual o superior a cero y permiten la flexibilidad necesaria para esos valores más atípicos (outliers).

A continuación, se definen las distribuciones asignadas a los parámetros en cuestión:

Parámetros	Distribución a Priori	Descripción
$\psi_1$	gamma (4, 4)	Ancho de Banda $2\ell^{2^{-1}}$ Año Desarrollo
$\psi_2$	gamma (4, 4)	Ancho de Banda $2\ell^{2^{-1}}$ Año Origen
$\eta^2$	$\mathcal{T}^+(4, 0, 1)$	Señal del GP
$\sigma^2$	$\mathcal{T}^+(4, 0, 1)$	Ruido del GP
$\alpha_1$	$\ln \mathcal{N}(0, 0.5)$	Alpha-Función Warping Año Desarrollo
$\beta_1$	$\ln \mathcal{N}(0, 0.5)$	Beta-Función Warping Año Origen
$\alpha_2$	$\ln \mathcal{N}(0, 0.5)$	Alpha-Función Warping Año Desarrollo
$\beta_2$	$\ln \mathcal{N}(0, 0.5)$	Beta -Función Warping Año Origen

*Tabla 2: Elaboración propia. Detalle con cada una de las distribuciones definidas a priori para cada uno de los parámetros en cuestión. (Fuente: Ruitenberg, (2019)).*

## V.II.III FUNCIONES DE COVARIANZA EMPLEADAS

Siguiendo con la metodología empleada por Lally y Hartman, (2018), a continuación visualizaremos una tabla con las funciones de covarianza seleccionadas finalmente para realizar la predicción.

Función Covarianza	Ecuación
Exponencial Cuadrático (SE)	$\eta^2(1 + \sqrt{3}d)\exp(-\sqrt{3}d)$
Matérn 3/2	$\eta^2(1 + \sqrt{5}d + 5/3d^2)\exp(-\sqrt{5}d)$
Matérn 5/2	$\eta^2\exp(-d^2)$

*Tabla 3: Elaboración propia. Funciones de Covarianza (Kernels) empleados para la predicción de provisiones, extrayendo los hiperparámetros a estimar. (Fuente: Ruitenberg, (2019)).*

En la tabla observamos las funciones de covarianza, pero no de la misma estructura que hemos definido inicialmente para los Kernels en este trabajo. Esto es debido a que la tabla 3 refleja las ecuaciones pero sin la integración de los hiperparámetros que debemos estimar. En esta tabla,  $d$  es igual a la distancia euclídea y  $d^2 = (x - x')^T \psi (x - x')$ , donde  $\psi$  = diagonal de  $(\psi_1, \psi_2)$  (en la tabla 2 se refleja las distribuciones a priori que seguirán en un inicio).

## V.IV.METODOLOGIA

El ejercicio consistirá en realizar diferentes predicciones sobre un mismo triangulo de provisiones. Por un lado, las obtenidas por el método de Chain-Ladder y por el otro lado, las obtenidas por los GP's empleados, diferenciando los outputs en base a los kernels seleccionados para cada predicción. La valoración de las mejores estimaciones se realizará a través de del error cuadrático medio de predicción (RMSE) acumulado.

$$RMSE \text{ agregado} = \sqrt{\frac{(\sum_{t=1}^T \hat{Z}_t - \sum_{t=1}^T Z_t)^2}{T}} \quad (56)$$

Donde  $T$  es el número total de observaciones de la muestra,  $Z_t$  el valor agregado que deseáramos obtener y  $\hat{Z}_t$  el valor agregado de las predicciones.

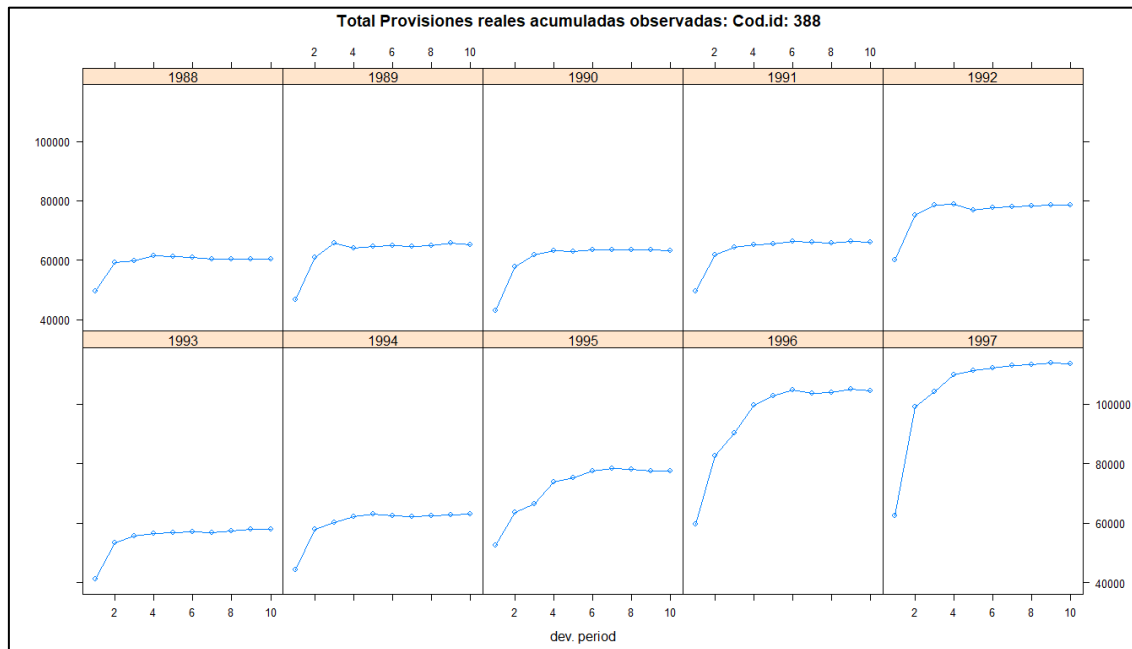
La estimación se realizará con un intervalo de confianza del 95% para la distribución posterior de las muestras obtenidas a través de caminos aleatorios multivariantes, en este caso, estimados por la metodología de Cadenas de Markov de Monte Carlo (MCMC). El método MCMC consiste en crear muestras autocorrelacionadas a partir de una variable aleatoria, con una densidad conocida, lo que nos permite obtener aproximaciones para la distribución posterior. Dicho de otro modo, en base a la distribución a priori ( $p(\theta)$ ) y el valor de la verosimilitud ( $p(x|\theta)$ ), el método devuelve la aproximación de una distribución posterior ( $p(\theta|x)$ ).

## V.V. BASES DE DATOS

La base de datos seleccionada pertenece a la librería pública de R “*casdata*”<sup>29</sup>, donde refleja triángulos enteros, desde el 1988 al 1997, obteniendo una matriz de 10x10, pudiendo obtener predicciones hasta el año 2006. En esta librería hay datos de múltiples empresas aseguradoras americanas para realizar ejercicios de predicciones con datos

<sup>29</sup> En el script de este trabajo se encuentra el desarrollo para instalar esta librería y filtrar la base de datos deseada.

reales observados. Para este trabajo se ha seleccionado “*comaauto*”, contiene datos de reclamaciones incurridas del ramo de automóviles, seleccionando la compañía del grupo Celina Mut (Id: 353)<sup>30</sup> y el grupo Federal Ins Co (id: 388), de este modo, podemos observar cómo se adaptan estos modelos en base a dos triángulos de provisiones diferentes. A continuación realizamos un vistazo inicial a los datos que contiene esta compañía con el conjunto de datos reales observados:



*Figura 8: Elaboración propia con RStudio. Visualización del total de provisiones acumuladas por cada uno de los años de origen del grupo Federal Ins Co.*

En la figura 8 podemos observar los datos reales de todas las provisiones por cada año de origen de la compañía en cuestión. Inicialmente podemos observar que el análisis de año de origen por sus diez años de desarrollo, mantienen un comportamiento heterogéneo, lo que puede dificultar nuestras predicciones, sobre todo para los últimos años de origen, donde se limita nuestro número de observaciones. A destacar el año 1988, con una caída en el tercer año de desarrollo, el 1994 como la que presenta un mayor crecimiento exponencial y por último, 1995, con un codo totalmente rígido.

## V.II.VI EXPLICACIÓN CÓDIGO STAN (INFERENCIA BAYESIANA)

Previo a visualizar los resultados, en este apartado explicaremos a modo resumen como funciona el código utilizado para la obtención de nuestras distribuciones a prior y a posterior a través del software estadístico Stan, perteneciente a Lally y Hartman, (2018). El código se estructura a través de los siguientes bloques:

En primer lugar, se crea una función, donde se definirá la función de la matriz de covarianza o Kernel escogido y se definirán los pesos de los inputs en base a la distribución seleccionada, en este caso, tal y como hemos comentado previamente, seguirán una función de distribución acumulada beta. Stan tiene predeterminadamente la

<sup>30</sup> Por motivos de sostenibilidad, solo visualizaremos el triángulo perteneciente a Federal Ins Co, la gráfica de Celina Mut puede encontrarse en el script de esta actividad.

función “*beta\_cdf*” que nos facilitará este cálculo<sup>31</sup>. Por otro lado, si que deberemos crear nuestra matriz Kernel personalizada en base a la forma predeterminada de la covarianza que hayamos seleccionado. Para obtener la distribución a prior, la matriz inversa se ha obtenido a través de la descomposición de Choleski, la cual, consiste en la factorización de la matriz en el producto de dos o más matrices, dicho de otro modo, crea una matriz auxiliar con las mismas dimensiones que la matriz de interés compuesta con ceros por encima de la diagonal principal multiplicada por su matriz traspuesta con ceros por debajo de la diagonal principal. Para nuestro caso, que nos encontramos frente a una matriz simétrica definida positiva, y tomando el input como un vector de variables tratadas (transformación ya comentada), esta puede ser factorizada por medio de una matriz triangular inferior y una matriz triangular superior:  $A = LL^T$ , siendo  $A$  nuestra matriz,  $L$  es la matriz triangular positiva y  $L^T$  la traspuesta.

En segundo lugar es el apartado de data. En este bloque se leerán nuestros datos que introduciremos en formato lista (“*list*”) para una correcta lectura de estos. Incluiremos nuestros inputs más los valores que nos definirán el tamaño dimensional de las predicciones que queremos introducir.

En tercer lugar, tenemos al apartado de transformación de datos, donde definiremos la media y la covarianza de nuestras distribuciones. Para esta actividad, no hemos integrado el ruido en nuestras predicciones, con lo cual, el vector media es cero para nuestras distribuciones a priori.

El cuarto bloque es para definir los hiperparámetros. En nuestro caso, ya hemos comentado previamente que no son valores fijos, sino que siguen unas distribuciones, por este motivo, definiremos para cada hiperparámetros la función que seguirá a priori (Tabla 2).

El quinto bloque es la definición del modelo y output. En este bloque se definen los parámetros fijos que siguen las funciones de nuestros hiperparámetros y el output deseado, en nuestro caso, una función que sigue la distribución multivariante de una normal a través de la descomposición de Choleski.

La principal ventaja que nos facilita realizar el código a través de Stan es que este implementa la probabilidad marginal, procedente de la inferencia bayesiana en su sistema.<sup>32</sup>

Por último, con la función “*stan*”, del paquete “*rstan*”, definiremos el escenario que queremos estructurar en nuestro modelo. La función tiene varios argumentos:

- Warmup: Número de iteraciones que no se usarán en la distribución posterior pero si servirán como un proceso de “calentamiento” del modelo. Definir un número de iteraciones de calentamiento es necesario para poder excluir esa porción de observaciones que se encuentran bajo el efecto de una posición inicial.
- Chains: El número de cadenas a definir de iteraciones. Estas cadenas son las cadenas de Markov, las cuales nos permitirán estimar la probabilidad en el siguiente período. Predeterminadamente toma un valor de 4 cadenas.

---

<sup>31</sup> Para un mayor detalle, pueden consultar la siguiente guía de usuario online de funciones de Stan: [https://mc-stan.org/docs/2\\_29/functions-reference/index.html](https://mc-stan.org/docs/2_29/functions-reference/index.html)

<sup>32</sup> Para mayor información, pueden consultar la siguiente guía de uso de Stan. El punto diez detalla cómo aplicar un GP en Stan: [https://mc-stan.org/docs/2\\_29/stan-users-guide/gaussian-processes.html](https://mc-stan.org/docs/2_29/stan-users-guide/gaussian-processes.html)

- Iter: Número de iteraciones que queremos aplicar a cada cadena. Definimos 2.000 iteraciones, de las cuales, solo 1.000 se usarán en el modelo al usarlas como Warmup. Nos quedaran 1.000 iteraciones para cada cadena, en total 4.000 iteraciones, que son las asignadas para la muestra de las distribuciones posteriores.

Este proceso nos devolverá la información de salida relevante para nuestro estudio, además, también nos devolverá otras salidas, analizadas en el script, pero que no visualizaremos en el trabajo. Seleccionaremos el promedio de los valores esperados de las iteraciones, el valor para los intervalos de confianza definidos, y el promedio de los valores de todos los parámetros definidos dentro del modelo.

## V.II.VII RESULTADOS PREDICCIONES PROVISIONES

<i>Cod.Id: 353</i>	<i>Obs</i>	<i>ChainLadder</i>	<i>GP-SE</i>	<i>GP-M.3/2</i>	<i>GP-M.5/2</i>
<i>IBNR</i>	4.272	3.125	3.420	3.493	3.743
<i>RMSE</i>	-	115	85	78	53

*Tabla 4: Elaboración propia. Cálculo del IBNR y del RMSE para los modelos en cuestión, en primer lugar Chain Ladder y seguido de los tres modelos estimados por GP, para la compañía Celina Mut.*

<i>Cod.Id: 388</i>	<i>Obs</i>	<i>ChainLadder</i>	<i>GP-SE</i>	<i>GP-M.3/2</i>	<i>GP-M.5/2</i>
<i>IBNR</i>	85.841	32.445	-6.693	39.291	53.706
<i>RMSE</i>	-	5.340	9.253	4.655	3.276

*Tabla 5: Elaboración propia. Cálculo del IBNR y del RMSE para los modelos en cuestión, en primer lugar Chain Ladder y seguido de los tres modelos estimados por GP, para la compañía Federal Ins Co.*

En la tabla 4 y 5 podemos observar los datos de salida estimador del cálculo del IBNR y del RMSE para los dos triángulos de provisiones estimados. En primer lugar, podemos destacar que aunque inicialmente los dos triángulos tienen comportamientos asimétricos, el modelo que mejor se ha ajustado a los datos reales es el GP estimado con un Kernel Matérn 5/2.

Por otro lado, el método Chain Ladder es el peor modelo que estima las predicciones para Celina Mut, aunque cabe decir, que para la compañía Federal Ins Co, su estimación es mejor que el GP Exponencial, el cual, ofrece un IBNR negativo, descartándolo directamente como método válido para estas dos casuísticas. Añadir que, aunque en Lally y Hartman, (2018) y Ruitenberg, (2019), el Kernel exponencial tuviese una muy buena adaptación, para nuestros datos no se ajusta debido a la tendencia individual de los años de desarrollo, pues siempre fuerza las predicciones a futuro hacia la media cero al no integrar ruido y con las distribuciones definidas a priori actuales. Aún así, podemos observar que todos los modelos retornan valores bastante alejados respecto al real observado.

Por motivos de sostenibilidad, solo visualizaremos el output gráfico del modelo que mejor se adapta a las predicciones finales.<sup>33</sup>

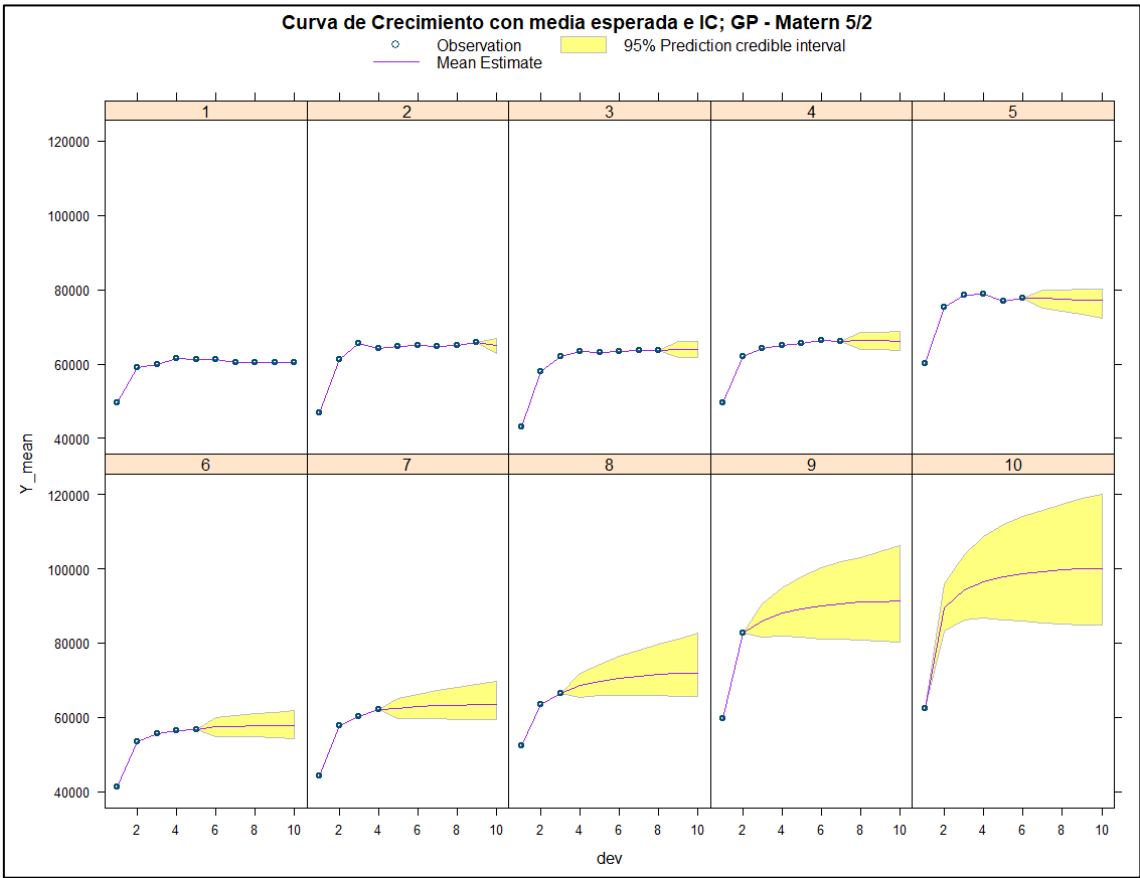


Figura 9: Elaboración propia con RStudio. Visualización de las predicciones del total de provisiones acumuladas por cada uno de los años de origen del grupo Federal Ins Co, estimado por GP – Kernel Matérn 5/2.

Si nos centramos en las diferencias entre la figura 9 y la figura 8 podemos extraer los siguiente: La predicción sigue un comportamiento bastante constante, al igual que los datos originales, lo cual facilita la predicción sin ruido en este Kernel, por otro lado, también podemos observar que la falta de importe total se debe en su mayoría a la predicción de los dos últimos años de origen, donde la predicción se queda corta en media, aunque sí que lo recoge en sus intervalos de confianza.

Cod.Id: 388	1	2	3	4	5	6	7	8	9	10
Obs	0	-486	-437	-110	871	1.043	870	10.834	22.058	51.198
GP-M.5/2	0	-842	192	126	-644	1.101	1.314	5.431	8.749	37.649

Tabla 6: Elaboración propia. Cálculo del IBNR por año de origen para el grupo Federal Ins Co.

<sup>33</sup> En el Anexo, script, incluye un análisis estadístico más detallado y la validación gráfica para cada uno de los modelos y grupo asegurador, con mayor detalle. También se encuentra un breve análisis de los parámetros, donde se refleja la media que toman con las diferentes iteraciones por MCMC, aunque cabe decir que toman valores relativamente constantes en común para las iteraciones planteadas en las cuatro cadenas.

Por último, si entramos en detalle, en la tabla 6 podemos observar el IBNR por año de predicción, observado grandes diferencias en los primeros próximos cuatro años y a partir del octavo año en adelante, donde el GP Matérn 5/2 se queda significativamente por debajo de los valores reales observados.



## VI CAPITULO IV – RESUMEN

### VI.I CONCLUSIONES

En esta actividad hemos explicado que es un GP en términos de regresión, como funciona y dos aplicaciones actuariales que pueden utilizarse. Aunque matemáticamente pueda parecer complejo, con RStudio y Stan hemos podido integrarlo en dos ramos diferentes. Por un lado, en seguros TAR como un ejercicio de interpolación, donde hemos observado parte del potencial de este GP, alcanzado el objetivo planteado inicialmente. La interpolación ha sido exitosa aun sin la realización de la optimización de los hiperparámetros. Por otro lado, también hemos integrado el GP para el cálculo de un monto de provisiones en No Vida, para el ramo automovilístico y para un el conjunto de ramos de una aseguradora, siguiendo la misma metodología aplicada en Lally y Hartman, (2018). Hemos alcanzado a observar que los diferentes tipos de GP son buenos predictores, en estas dos casuísticas, se adaptan mejor que el método convencional de Chain Ladder, reflejando el potencial de este. Aun así, cabe destacar la importancia de la selección del Kernel y las distribuciones a priori de los parámetros, pues hemos observado grandes diferencias en base a la selección de este, destacando el Kernel Exponencial como el que mejor ajusta en el ejercicio de interpolación y el Kernel Matérn 5/2 como el que mejor se ajusta en el ejercicio de cálculo del IBNR.

En base a la experiencia obtenida, concluimos en que el GP tiene un gran potencial, siempre y cuando se haya realizado un correcto proceso de entrenamiento, destacando su fuerza en términos de interpolación pero observando debilidades en predicciones a futuro, al menos, con las distribuciones a priori que hemos definido en este ejercicio.

Por último, concluir que la integración de un GP, sin el uso de librerías específicas<sup>34</sup> tiene un alto coste computacional y nos hemos visto frente algunas limitaciones. Por ejemplo, el ejercicio de interpolación mensual, en un inicio se buscaba la actualización diaria, pero debido a estos problemas, se ha reducido al tratamiento mensual. Para solucionar este problema, puede realizarse un GP para cada par de puntos de la muestra, es decir, en vez de extraer la continuidad de toda la regresión, puede realizarse por tramos, permitiendo optimizar este coste computacional. Por otro lado, ejecutar la inferencia bayesiana en Rstan, supone un tiempo de espera entre 10 y 20 minutos por cada modelo estimado.<sup>35</sup>

### VI.II DEBATE Y PRÓXIMOS PASOS

Tras las conclusiones obtenidas, en este apartado detallaremos cuales podrían ser las posibles mejoras y discusiones de este trabajo y próximos pasos que se podrían realizar.

En primer lugar, respecto la aplicabilidad empírica de provisiones, podríamos valorar la integración de ruido en nuestras distribuciones con la finalidad de obtener una mayor flexibilidad. Además, podría contemplarse aplicar otras distribuciones a priori para nuestros parámetros. También se podría realizar estas predicciones sobre triángulos más

---

<sup>34</sup> En este ejercicio hemos tratado de usar el mínimo de librerías posibles y tratar de desarrollar el GP por pasos, pero en el entorno de R, existen varias librerías que nos permitirían obtener una resolución más rápida y/o eficiente. Algunas de las más relevantes: “GauPro”, “GPfit”, “varycoef” y “GPFDA”. También comentar que en Python hay todo un desarrollo sobre GP’s.

<sup>35</sup> Trabajo realizado desde un ordenador portátil, Asus (Procesador: Intel(R) Core(TM) i7-9th Gen), RAM de 16 GB y 1.000 GB de disco duro.

recientes, pues recordemos que sea realizado sobre dos triángulos comprendidos entre el 1988 al 1997, obteniendo las predicciones máximas hasta la anualidad 2006. Por último, sería interesante observar los datos de salida usando el monto de primas respecto cada año de origen como variable exógena a la hora de estimar, al igual que realizó Ruitenberg, (2019).

En segundo lugar, respecto a la aplicabilidad empírica sobre los seguros TAR, hemos observado que realmente podría ser una herramienta a integrarse debido a los buenos resultados obtenidos, pero también habría que observar sobre otras tablas de mortalidad y comparar si el Kernel Exponencial sigue siendo de utilidad. También cabe comentar que solo se han tratado los inputs de los valores de cohorte, estandarizados, pero podría valorarse tratar la introducción de los datos de entrada del vector de edades actuariales, similar al procedimiento que se ha realizado en el cálculo de provisiones, pero solo con un input, además de incluir el proceso de optimización de hiperparámetros.

Esta interpolación también puede ser de gran utilidad para el cálculo de primas cuando el período de cobertura es inferior a la anualidad y se solicita cobertura para meses, semanas o días, dando una mayor aproximación y beneficiando al asegurado, sin perder el criterio actuarial como aseguradora que garantice su solvencia frente a un siniestro. Además, como las tablas de mortalidad llevan un período de actualización de diez años, podría plantearse la integración de esta metodología en este y otros ámbitos, consiguiendo una mayor competitividad para la aseguradora sin indagar en un mayor riesgo, siempre y cuando se consiguiese realmente una buena estimación del GP.

Respecto a los dos ejemplos empíricos, también cabe decir que tanto RStudio como Stan permiten obtener datos de salidas cuando las variables son No-Gaussianas a través de la distribución de Bernoulli con parámetros Logit, una posible línea de investigación futura interesante.

Por otro lado, sin limitarnos a estas dos aplicaciones, la estimación del GP realmente tiene varias aplicaciones en ámbitos de ingeniería, matemáticas, salud y medioambiental, donde, tras una lectura bibliográfica, hay varias materias que nos permitirían obtener nuevas aplicaciones y mejores estimaciones en el ámbito actuarial. Algunos de ellos son los Kernels Deep Learning (Kernels de aprendizaje automático), comparativa del GP con máquinas de soporte vectorial (SVM), las cuales se clasifican como algoritmos de aprendizaje automático, la implementación de Kernels compuestos (nuevas funciones Kernels que son la composición conjunta de otros Kernels ya predefinidos) y por último, profundizar en los GP para árboles de clasificación.

## VI.III BIBLIOGRAFÍA

- Alexis Sáez, R. (2016). *La estimación de provisiones técnicas y el efecto calendario*. Universidad de Granada.
- Andrew Gordon Wilson, David A. Knowles, & Zoubin Ghahramani (2012). *Gaussian Process Regression Networks*. In ICML.
- Ayuso. M, Corrales. H, Guillén. M, Pérez-Marín. A.M, Rojo. J.L. (2007). *Estadística actuarial vida*. Barcelona: Universitat de Barcelona.
- Boettiger, C. (2012). *Basic regression in Gaussian processes*. <https://www.carlboettiger.info/2012/10/17/basic-regression-in-gaussian-processes.html>
- Clark (2020, Nov. 30). Michael Clark: *Models by Example*. Retrieved from <https://m-clark.github.io/posts/2020-11-30-models-by-example/>
- Flaxman, S., Gelman, A., Neill, D.B., Smola, A., & Vehtari, A. (2015). *Fast hierarchical Gaussian processes*. (Unpublished manuscript).
- Jackman, S.: “*Bayesian Analysis for Social Science*” Willey, 2009.
- James Guszcza. “Hierarchical growth curve models for loss reserving”. In: *Casualty Actuarial Society Forum*. 2008
- Jie Wang. (2020). *An Intuitive Tutorial to Gaussian Processes Regression*.
- Lally, N. y Hartman, B. (2018). Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping. *Insurance: Mathematics and Economics*, 82, 124–140.
- Lincoln, C., Tobar, F. (2021). *The Art of Gaussian Processes: Classical and Contemporary*.
- Ludkovski, Mike, and Howard Zail. 2021. “*Gaussian Process Models for Incremental Loss Ratios*.” *Variance* 15 (1)
- Rosen, J. y Marcus, M. B. (2011). *Markov Processes, Gaussian Processes, and Local Times*. Cambridge University Press.
- Peng Shi, Sanjib Basu, and Glenn G Meyers. *A Bayesian log-normal model for multivariate loss reserving*. In: *North American Actuarial Journal* 16.1 (2012).
- Rasmussen, C. E., Williams, C. K. I. y Bach, F. (2005). *Gaussian Processes for Machine Learning*. MIT Press.
- Robert B. Gramacy (2020). *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC.

- Ruitenberg, P. (2019). *Adapting a Hierarchical Gaussian Process model to predict the loss reserve of a non-life insurer.*, University of Twente.
- Tobar, F. (2020). *Aprendizaje de Máquinas*. Universidad de Chile. <https://www.studocu.com/cl/document/universidad-de-chile/deep-learning/aprendizaje-de-maquinas/11400390>
- Shi, J. Q. (2011). *Gaussian process regression analysis for functional data*. CRC Press.
- Shi, P., Basu, S. y Meyers, G. G. (2012). A Bayesian Log-Normal Model for Multivariate Loss Reserving. *North American Actuarial Journal*, 16(1), 29–51. <https://doi.org/10.1080/10920277.2012.10590631>
- Snoek, J., Swersky, K., Zemel, R.S., & Adams, R.P. (2014). *Input Warping for Bayesian Optimization of Non-Stationary Functions*. ICML.
- Suwandani, R. N. y Purwono, Y. (2021). Implementation of Gaussian Process Regression in Estimating Motor Vehicle Insurance Claims Reserves. *Journal of Asian Multicultural Research for Economy and Management Study*, 2(1), 38–48. <https://doi.org/10.47616/jamrems.v2i1.77>

## VI.IV ANEXOS

Por criterios de sostenibilidad, el código utilizado para la realización de esta actividad se encuentra en el enlace adjunto a continuación. Además, también incluye algunos comentarios relevantes para esos datos que no se han incluido en este trabajo pero si pueden considerarse de interés.

- Script: [https://github.com/drius1996/TFM\\_GPR](https://github.com/drius1996/TFM_GPR)

### Índice Tablas.

- *Tabla 1: Elaboración propia. Comparativa entre las primas netas calculadas por la metodología tradicional contra la nueva metodología estimada por GP. Edad y PN son las columnas calculadas por la metodología tradicional y Edad'' y PN''(INT) por la nueva metodología. .... 32*
- *Tabla 2: Elaboración propia. Detalle con cada una de las distribuciones definidas a priori para cada uno de los parámetros en cuestión. (Fuente: Ruitenberg, (2019)). .... 34*
- *Tabla 3: Elaboración propia. Funciones de Covarianza (Kernels) empleados para la predicción de provisiones, extrayendo los hiperparámetros a estimar. (Fuente: Ruitenberg, (2019)). .... 35*
- *Tabla 4: Elaboración propia. Cálculo del IBNR y del RMSE para los modelos en cuestión, en primer lugar Chain Ladder y seguido de los tres modelos estimados por GP, para la compañía Celina Mut. .... 38*
- *Tabla 5: Elaboración propia. Cálculo del IBNR y del RMSE para los modelos en cuestión, en primer lugar Chain Ladder y seguido de los tres modelos estimados por GP, para la compañía Federal Ins Co. .... 38*
- *Tabla 6: Elaboración propia. Cálculo del IBNR por año de origen para el grupo Federal Ins Co. .... 39*

### Índice Figuras.

- *Figura 1: Elaboración propia con RStudio. Inferencia bayesiana para la obtención de una distribución a posteriori adaptada a los datos de entrada. .... 10*
- *Figura 2: Elaboración propia con RStudio. Distribuciones a priori de los siguientes kernels en base a la modificación del parámetro length-scale: Exponencial, Mattern 3/2 y Mattern 5/2. .... 15*
- *Figura 3: Elaboración propia con RStudio. Distribuciones a priori de los siguientes kernels en base a la modificación del parámetro length-scale, alpha y periodicidad: Racional Cuadrático y Periódico. .... 15*
- *Figura 4: Elaboración propia con RStudio. Visualización de una regresión por proceso gaussiano, con Kernel exponencial, ancho de banda definido igual a 0.1 y ruido 0.2. El intervalo de confianza asignado tiene un nivel de confianza asignado del 95% (bandas grises). .... 19*

- *Figura 5: Elaboración propia con RStudio. Valor de cohorte comprendido entre la edad de 0 años a 120 años, para una muestra de 1.000.000 personas, perteneciente a la tablas PASEM Unisex 2020 de 2no orden. .... 29*
- *Figura 6: Elaboración propia con RStudio. Interpolación del Valor de cohorte comprendido entre la edad de 0 años a 120 años, para una muestra de 1.000.000, perteneciente a la tablas PASEM Unisex 2020 de 2no orden, con Kernel Exponencial y parámetro ancho de banda constante igual a 1 y valor sigma igual a 0.15. .... 30*
- *Figura 7: Elaboración propia con RStudio. Comparativa entre las probabilidades de fallecimiento anuales contra las probabilidades de fallecimiento interpoladas y tratadas para el tramo de 54 a 60 años. .... 31*
- *Figura 8: Elaboración propia con RStudio. Visualización del total de provisiones acumuladas por cada uno de los años de origen del grupo Federal Ins Co. .... 36*
- *Figura 9: Elaboración propia con RStudio. Visualización de las predicciones del total de provisiones acumuladas por cada uno de los años de origen del grupo Federal Ins Co, estimado por GP – Kernel Matérn 5/2. .... 39*

### Índice Triángulos.

- *Triangulo 1: Elaboración propia. Plantilla de un cuadro de provisiones técnicas. .... 26*