

Técnicas de aprendizaje automático aplicadas a la estimación del Valor a Riesgo y otras medidas de riesgo financiero



Hugo Naudín López
Trabajo de fin de grado de Matemáticas
Universidad de Zaragoza

Junio de 2025

Resumen

Este documento es una plantilla en \LaTeX diseñada para cumplir con las *directrices propias* para la elaboración del trabajo de fin de grado de Matemáticas en la Facultad de Ciencias de la Universidad de Zaragoza. Estas directrices, que se pueden ver en [8], indican lo siguiente sobre la memoria (artículo 4):

1. La memoria del TFG podrá ser redactada en español o en inglés. Si la memoria está escrita en español (respectivamente en inglés) deberá acompañarse de un resumen en inglés (respectivamente en español). La extensión de dicho resumen será de dos o tres páginas.
2. La memoria se editará, preferiblemente, utilizando el sistema de composición de documentos \LaTeX . A tal efecto, en la página web de la Facultad de Ciencias se dispondrá de una plantilla \LaTeX .
3. Con el fin de facilitar su validación administrativa, todas las memorias de TFG de la Facultad de Ciencias tendrán un tamaño de letra diferenciado mínimo de 11 puntos, con un interlineado a espacio 1,15 y márgenes de al menos 2,5 cm, en papel tamaño DIN A4. El índice y el resumen deberán ir justo antes del inicio de la memoria y esta, en ningún caso, podrá superar las 25 páginas, excluidos los anexos.
4. El director de un TFG deberá comprobar que la memoria presentada cumple con las normas de edición establecidas, antes de dar el visto bueno a su depósito.

Esta plantilla se ofrece por si resulta útil, pero debe quedar claro que no es obligatorio usarla: el trabajo se puede escribir prescindiendo de esta plantilla, siempre que se respeten las directrices anteriores.

La plantilla está diseñada para un trabajo compuesto por uno o más capítulos y una bibliografía única, es decir, no una por cada capítulo. Detallando un poco más, la estructura es la siguiente y por este orden:

- Una portada.
- Un resumen.
- Después, un índice general en el que figuran el resumen, los capítulos de los que consta el trabajo y la bibliografía. Cada capítulo puede contener secciones, que también aparecen en el índice.
- La parte principal, formada por cada uno de los capítulos que componen el trabajo.
- Y, por último, la bibliografía.

La plantilla está escrita para componerse en pdf\LaTeX y la forman varios archivos que deben estar en la misma carpeta:

- Un documento principal, que contiene las órdenes iniciales y el código \LaTeX del texto.
- Los archivos de imágenes que se vayan a incluir en el trabajo.

En el primer capítulo de este modelo se dan más detalles sobre esta plantilla.

Obsérvese que en este ejemplo el resumen no sigue las directrices establecidas para los trabajos de fin de grado, según las cuales debe estar escrito en inglés si el trabajo está en castellano.

El último capítulo de este documento se ha puesto solo como relleno, como ejemplo de texto matemático, y no corresponde a un verdadero trabajo de fin de grado.

Índice general

Resumen	III
1. Introducción: el riesgo en las coporaciones	1
1.1. Valor a Riesgo (VaR)	1
1.1.1. Ventajas y desventajas del VaR	2
1.2. Expected Shorfall (ES) o Valor a Riesgo Condicional (CVaR)	2
1.2.1. Ventajas y desventajas del ES	2
1.3. Interés de las medidas de riesgo	3
1.3.1. VaR y ES orientado a carteras de valores	3
1.3.2. VaR y ES orientado a inversiones inmobiliarias	3
1.3.3. VaR y ES orientadas a corporaciones	4
1.3.4. VaR y ES orientado a predicciones de impago en créditos	4
2. Modelos de riesgo	5
2.1. Modelos tradicionales	5
2.1.1. Regresión cuantil	5
2.1.2. Regresión expectil	7
2.2. Modelos basados en Machine Learning	9
2.2.1. Redes Neuronales	9
2.2.2. Algoritmos de Boosting	13
3. Aplicaciones de modelos a datos de créditos	15
4. Medidas de riesgo	17
4.1. VaR	17
4.2. CoVar	18
4.3. Déficit esperado (Expected Shotfall)	18
4.4. Modelos de riesgo	19
4.4.1. Modelos clásicos	20
4.4.2. Modelo GARCH	20
4.4.3. Modelos de aprendizaje automático	22
4.4.4. Modelo XGBoost	22
5. Aplicación de QR y XGBoost a ...	27
6. Análisis de resultados	29
7. Conclusión	31
Bibliografía	33

Capítulo 1

Introducción: el riesgo en las coporaciones

A la hora de invertir, el interés de las corporaciones financieras es maximizar el beneficio y minimizar las pérdidas. Para ello, las compañías utilizan herramientas entre las que se encuentra la gestión del riesgo. El riesgo es la incertidumbre sobre la evolución de un activo, e indica la posibilidad de que una inversión ofrezca un rendimiento distinto del esperado, por lo que la gestión y el control de él puede ayudar a minimizar las pérdidas y optimizar las ganancias.

Durante la década de los 80, la Comisión de Bolsa y Valores de los Estados Unidos (SEC) comenzó a divulgar cómo las empresas estaban expuestas a la volatilidad de los mercados financieros, introduciendo y dando gran importancia al concepto de riesgo en el sistema, debido a los acontecimientos vividos en aquel entonces (Lunes Negro en 1987 y la Crisis de Ahorros y Prestamos). Más adelante, a partir de la crisis financiera del año 2008 y el acuerdo de Basilea III, las políticas del riesgo que podían correr los bancos a la hora de conceder créditos y realizar inversiones se endurecieron, lo que hizo que los modelos de predicción y gestión de riesgo se optimizaran y mejoraran de manera considerable.

Por otra parte, el uso de la inteligencia artificial (IA) y los algoritmos basados en el aprendizaje automático han incrementado en los últimos años, llegando a introducirse también en el campo financiero y en la implementación de modelos alternativos a la hora de gestionar el riesgo.

Para medir el riesgo, existen una gran cantidad de métodos y medidas utilizadas globalmente para ayudar a las entidades financieras a tomar decisiones. Entre ellas, destaca el Valor a Riesgo

1.1. Valor a Riesgo (VaR)

Se define el Valor a Riesgo (VaR) como la cuantía máxima de dinero que puede perderse en un período para un nivel específico de confianza. Es decir, si definimos G_t como los beneficios que obtenemos de una inversión en un tiempo t (pueden ser negativos), entonces:

$$\mathbb{P}\{G_t \leq VaR_{\alpha,t}\} = 1 - \alpha$$

con α un nivel de confianza específico. Por otro lado, tomamos unas variables aleatorias que simbolizan las ganancias y pérdidas de una inversión en un tiempo t , G_t , las cuales siguen una distribución F . Entonces, podemos definir el VaR como:

$$VaR_{\alpha,t} = \inf\{x \in \mathbb{R} : F(x) \geq 1 - \alpha\}$$

Es decir, dados unos datos empiricos de ganancias (las pérdidas son negativas), el VaR representa el cuantil α de la distribución empírica, y por tanto:

$$VaR_{\alpha,t} = F^{-1}(1 - \alpha)$$

1.1.1. Ventajas y desventajas del VaR

La medida VaR tiene una gran utilidad a la hora de limitar el riesgo y controlar las potenciales pérdidas (o retornos inferiores a lo esperado) de una inversión, pero hay otros aspectos en los que esta medida no es completa.

Por una parte, es una medida muy simple e intuitiva. Su significado y lo que representa son fácilmente comprensible, lo que hace que su utilización este muy extendida. Además, la comparabilidad de dos distribuciones se puede realizar fácilmente mediante esta medida, lo que puede resultar útil para evaluar diferentes carteras o estrategias de inversión en términos de su riesgo potencial. Por ejemplo, al comparar dos activos, el VaR permite identificar cuál de ellos tiene una mayor probabilidad de incurrir en pérdidas superiores a un cierto umbral en un horizonte de tiempo definido. Esto facilita la toma de decisiones por parte de gestores de riesgo e inversionistas. Finalmente, hay en ocasiones en las que interesa la estabilidad de las medidas de riesgo. A la hora de hallar el VaR de una distribución, cualquier cuantil inferior al $1 - \alpha$ puede tener cualquier valor, pero el VaR no variará, ya que es independiente de estos valores, siempre y cuando sean menores.

Sin embargo, que el VaR no varíe aunque los cuantiles menores sí, puede ser negativo desde otro punto de vista. Esto significaría que una mínima variación del nivel de confianza significaría una gran variación del VaR. Por tanto, esta medida estaría ocultando un riesgo mayor al que parece en un principio.

Esta desventaja hace del VaR una medida que, a pesar de su gran utilidad, presenta una notable incompletitud. Por ello, existen otras medidas que complementan a esta primera, ofreciendo una extensión a la información del VaR, sobre todo, dando información del cuantil inferior. Entre estas variables se encuentra el Expected Shortfall (ES) o Valor a Riesgo Condicional (CVaR).

1.2. Expected Shortfall (ES) o Valor a Riesgo Condicional (CVaR)

El Expected Shortfall (ES) o Valor a Riesgo Condicional (CVaR) son dos términos relacionados con el mismo concepto. Para no sobrecargar, se referirá a este concepto como ES de aquí en adelante. El ES es una medida complementaria al VaR que se define como la pérdida esperada en una inversión sabiendo que la pérdida es mayor que el valor del VaR. Por tanto, dado G_t un beneficio de una inversión en el tiempo t , formalmente definimos:

$$ES_{\alpha,t} = \mathbb{E}(G_t | G_t \leq VaR_{\alpha,t})$$

El ES siempre será una beneficio menor al representado por el VaR, ya que el ES toma en cuenta los beneficios menores a él y, por lo, tanto refleja una evaluación más conservadora del riesgo. Si el VaR se considera como el valor máximo de pérdida en un nivel de confianza α , el ES estima el valor promedio de las pérdidas cuando las pérdidas ya superan ese valor. Por lo tanto, el ES proporciona una medida más completa del riesgo asociado a eventos extremos.

1.2.1. Ventajas y desventajas del ES

La medida ES, al contrario del VaR, es una medida continua respecto a α , es decir, cambios pequeños en el nivel de confianza no suponen cambios grandes en el CVaR, que describe como es la pérdida esperada extrema, en caso de haberla. Si tomamos una sucesión de inversiones $(X_i)_{i=1}^n$ y un valor asociado a cada una de ellas $(w_i)_{i=1}^n$, el cual representa la proporción que representa esa inversión en el total de nuestra cartera ($\sum_{i=1}^n w_i = 1$), entonces la medida ES en un tiempo t con una confianza α es convexa respecto a la combinación de estas variables:

$$ES_{\alpha,t}(w_1 X_1 + \dots w_n X_n) \leq w_1 ES_{\alpha,t}(X_1) + \dots w_n ES_{\alpha,t}(X_n)$$

lo cual es coherente con la idea de la diversificación de activos para la reducción del riesgo, ya que el ES de la cartera nunca excede el promedio ponderado de los ES individuales

Por otro lado, esta medida es muy sensible a los errores de estimación. Normalmente, no existen muchos datos de los cuantiles bajos, por lo que es necesario realizar un modelo para obtener más datos. Existen casos en los que estos modelos son erráticos y los datos en las colas pueden desviarse de la realidad.

Por tanto, la utilización y corrección de esta medida depende en gran parte de la corrección y confianza que tengamos en los datos. Sin embargo, la combinación de esta medida con el VaR ofrece una visión muy completa del riesgo en la inversión.

1.3. Interés de las medidas de riesgo

Las medidas de riesgo, como el Valor en Riesgo (VaR) y Expected Shortfall (ES), son herramientas fundamentales en la gestión de riesgos, ya que permiten cuantificar y gestionar la incertidumbre asociada a las pérdidas potenciales en diferentes contextos. Su utilidad varía dependiendo de si se aplican a carteras de valores o al ámbito corporativo.

1.3.1. VaR y ES orientado a carteras de valores

En el ámbito de la gestión de carteras de inversión, el VaR y el ES son cruciales para entender y controlar la exposición al riesgo.

En el caso del VaR, estima la pérdida máxima de una cartera de valores en un horizonte temporal específico y con un nivel de confianza dado. Permite a los gestores de carteras evaluar riesgos acumulativos, identificando el impacto potencial de eventos de mercado extremos sobre la cartera, tomar decisiones informadas, comparando diferentes estrategias de inversión en función de su perfil de riesgo y cumplir requisitos regulatorios (normativas de Basilea).

El ES complementa al VaR al estimar la pérdida promedio en los casos en que esta exceda el VaR. Es especialmente útil porque ofrece una visión más completa y optimiza carteras.

Las dos medidas juntas son esenciales para medir y gestionar el riesgo de mercado, mejorar la asignación de activos y proteger a los inversores frente a pérdidas inesperadas.

1.3.2. VaR y ES orientado a inversiones inmobiliarias

La utilización de estas medidas de riesgo en las inversiones inmobiliarias se enfoca en la predicción del precio de los alquileres de los inmuebles. Dado el precio de una vivienda, el VaR y el ES estiman escenarios negativos de precios a los que puede ser alquilado, permitiendo estimar si la inversión es suficientemente rentable.

Al mismo tiempo, el VaR y el ES se pueden utilizar para calcular una cota inferior del valor intrínseco de una vivienda dadas unas características específicas, como la ubicación, el tamaño y las condiciones del mercado. Esto permite a los inversores evaluar si el precio de compra de una propiedad es razonable en relación con su valor potencial de alquiler y su riesgo asociado.

1.3.3. VaR y ES orientadas a corporaciones

En el contexto corporativo, estas métricas ayudan a evaluar el riesgo financiero y operativo que enfrentan las empresas en sus actividades.

El VaR se utiliza para medir riesgos asociados a las fluctuaciones en tipos de cambio, tasas de interés o precios de materias primas y a la toma de decisiones estratégicas.

Por su parte el ES proporciona una perspectiva más amplia, considerando las pérdidas extremas que pueden no ser capturadas por el VaR. Ofrece un análisis más detallado de los escenarios de mayor impacto financiero y permite a las empresas diseñar planes de mitigación y estrategias para afrontar eventos inesperados.

En el ámbito corporativo, estas medidas de riesgo facilitan la identificación de vulnerabilidades clave, la planificación financiera y la protección de los recursos de la empresa frente a incertidumbres del entorno operativo y del mercado.

1.3.4. VaR y ES orientado a predicciones de impago en créditos

Estas variables económicas también se emplean en la predicción de impago de créditos, lo que ayuda a los bancos y entidades de prestamos en sus decisiones. El VaR se utiliza para estimar la cantidad que con una probabilidad dada el cliente impagará, considerando factores como el historial crediticio, la situación financiera, las condiciones del mercado y determinadas características del cliente.

Por otra parte, el ES, al igual que en el caso de las carteras de valores, se utiliza para evaluar la pérdida promedio en los casos en que el cliente no pague, lo que da más información a la entidad acerca de la rentabilidad de dar el crédito.

Estas medidas permiten a las entidades evaluar el riesgo asociado a la concesión de créditos y ajustar sus políticas de préstamo en consecuencia.

Capítulo 2

Modelos de riesgo

Con el objetivo de aproximar las medidas de riesgo anteriormente descritas, a lo largo de la historia se han ido desarrollando diferentes modelos, que emplean técnicas distintas, que ayudan a predecir el riesgo dadas distintas situaciones.

2.1. Modelos tradicionales

Los métodos tradicionales están fundamentados en la econometría y en métodos estadísticos. Sin embargo, estos enfoques pueden presentar limitaciones cuando se enfrentan a condiciones de mercado altamente volátiles o a eventos inesperados. Además, algunos se basan en suposiciones no siempre ciertas (distribuciones normales, independencia de errores, etc.). Sin embargo, existen otras técnicas que tienen una mayor robustez.

2.1.1. Regresión cuantil

La regresión cuantil es un método estadístico que permite estimar la relación entre variables explicativas con los cuantiles de la distribución de la variable dependiente. A diferencia de la regresión lineal usual, que se centra en la media condicional, la regresión cuantil proporciona una visión más completa de la relación entre variables al considerar los cuantiles de la distribución, de donde podemos obtener una información mayor.

Para introducir la regresión cuantil, veamos primero las bases de la regresión lineal. Tomamos unas variables explicativas $\mathbf{X}_i = \{X_{i1}, X_{i2}, \dots, X_{ik}\}$ y una variable dependiente Y_i . La regresión lineal se basa en la suposición de que existe una relación entre las variables explicativas y la variable dependiente y se puede expresar de la siguiente manera:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

con ε una variable de ruido que sigue una distribución $N(0, \sigma)$ dado un σ cualquiera. Es decir, existe un vector $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ tal que:

$$Y_i = \beta \cdot \mathbf{X}_i + \varepsilon_i$$

Los coeficientes β_i representan la relación entre las variables y se denominan coeficientes de regresión. El objetivo de la regresión lineal es obtener unos coeficientes $\hat{\beta}_{ik}$:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \dots + \hat{\beta}_k X_{ik}$$

tal que la suma de los errores al cuadrado sea mínima. Es decir, queremos obtener $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ tal que:

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

La regresión cuantil, por otro lado, intenta modelar el cuantil α de la variable dependiente Y_i dado un conjunto de variables explicativas \mathbf{X}_i , en vez de centrarse en la media. La forma general del modelo es:

$$c_\alpha(Y_i|\mathbf{X}_i) = \dot{\beta}_0(\alpha) + \dot{\beta}_1(\alpha)X_{i1} + \dot{\beta}_2(\alpha)X_{i2} + \cdots + \dot{\beta}_k(\alpha)X_{ik}$$

donde $c_\alpha(Y_i|\mathbf{X}_i)$ es la estimación del cuantil α de la variable dependiente Y_i dado el vector de variables explicativas \mathbf{X}_i .

Paralelamente a lo anterior, para hallar los coeficientes de regresión $\dot{\beta} = (\dot{\beta}_0(\alpha), \dot{\beta}_1(\alpha), \dots, \dot{\beta}_k(\alpha))$ debemos hallar los valores que minimicen una función de pérdida cuantílica. En este caso, esta función va ser diferente a la anterior:

$$\dot{\beta}(\alpha) = \operatorname{argmin}_{\beta} \sum_{i=1}^n \rho_\alpha(Y_i - c_\alpha(Y_i|\mathbf{X}_i))$$

con $\rho_\alpha(r) = r \cdot (\alpha - I(r < 0))$ la función de pérdida cuantílica.

Proposición 1. Sea X una variable aleatoria continua con función de distribución F y cuantil α , y dada la función de pérdida cuantílica:

$$\rho_\alpha(r) = r(\alpha - I(r < 0))$$

Tomando una variable θ y la función:

$$E_x(\rho_\alpha(X - \theta)) = \int \rho_\alpha(x - \theta) f_x(x) dx$$

entonces, el valor de θ que minimiza la función es el cuantil α de la variable aleatoria X .

Demostración. Nuestro objetivo es ver que:

$$c_\alpha(X) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\} = \operatorname{argmin}_{\theta} \{E_x(\rho_\alpha(X - \theta))\}$$

Para demostrarlo, vamos a ver que el valor de θ que minimiza la función $E_x(\rho_\alpha(X - \theta))$ es el cuantil α de la variable aleatoria X :

$$\begin{aligned} E_x(\rho_\alpha(X - \theta)) &= \int \rho_\alpha(x - \theta) f_x(x) dx \\ &= \int_{-\infty}^{\infty} (x - \theta)(\alpha - I(x - \theta < 0)) f_x(x) dx \\ &= \int_{-\infty}^{\theta} (x - \theta)(\alpha - 1) f_x(x) dx + \int_{\theta}^{\infty} (x - \theta) \alpha f_x(x) dx \\ &= (\alpha - 1) \int_{-\infty}^{\theta} (x - \theta) f_x(x) dx + \alpha \int_{\theta}^{\infty} (x - \theta) f_x(x) dx \\ &= (\alpha - 1) \int_{-\infty}^{\theta} x f_x(x) dx - (\alpha - 1) \theta \int_{-\infty}^{\theta} f_x(x) dx + \alpha \int_{\theta}^{\infty} x f_x(x) dx - \alpha \theta \int_{\theta}^{\infty} f_x(x) dx \end{aligned}$$

Como queremos minimizar esta función con respecto a θ , queremos que su derivada sea nula, por lo que derivando con respecto a θ :

$$\begin{aligned} 0 &= -(\alpha - 1) \int_{-\infty}^{\theta} f_x(x) dx - \alpha \int_{\theta}^{\infty} f_x(x) dx \\ 0 &= -(\alpha - 1)(F(\theta) - F(-\infty)) - \alpha(F(\infty) - F(\theta)) \\ 0 &= -(\alpha - 1)F(\theta) - \alpha(1 - F(\theta)) \\ 0 &= -\alpha F(\theta) + F(\theta) - \alpha + \alpha F(\theta) \\ 0 &= F(\theta) - \alpha \implies \theta = F^{-1}(\alpha) \end{aligned}$$

Por tanto, el valor de θ que minimiza la función es el cuantil α de la variable aleatoria X . □

Un vez estimamos los coeficientes, podremos hacer una predicción de los cuantiles de la variable dependiente Y_i dado un conjunto de variables explicativas \mathbf{X}_i .

La utilidad de la regresión cuantil reside, entre otras, en la facilidad para el cálculo del VaR, ya que permite estimar beneficios extremadamente bajos (incluso negativos) y relacionarlos con otros factores como las tasas de interés, los precios de las materias primas o las condiciones del mercado, todo ello sin depender de la media ni de supuestos de normalidad.

2.1.2. Regresión expectil

Además de los cuantiles, existe una medida estadística de gran utilidad para la estimación del riesgo. El expectil τ (e_τ) de una variable aleatoria X es aquel valor que minimiza la función $\phi(X - e_\tau)$, donde:

$$\phi(r) = \begin{cases} \tau \cdot r^2 & \text{si } r \geq 0 \\ (1 - \tau)r^2 & \text{si } r < 0 \end{cases}$$

Los expectiles se pueden utilizar como una aproximación de los cuantiles, aunque no son exactamente lo mismo. Mientras que los cuantiles dividen la distribución en partes iguales, los expectiles ponderan las pérdidas de manera diferente, dando más peso a pérdidas extremas. Por tanto, la aproximación no es la manera correcta de usarlos.

La regresión expectil es un método estadístico útil para estimar variables relacionadas con la cola de una distribución como, por ejemplo, el Expected Shortfall (ES). Mientras que la regresión cuantil minimiza una función de pérdida absoluta asimétrica, la regresión expectil minimiza una función de pérdida cuadrática asimétrica.

Dadas unas variables explicativas $\mathbf{X}_i = \{X_{i1}, X_{i2}, \dots, X_{ik}\}$ y una variable dependiente Y_i , la regresión expectil se define como:

$$e_\tau(Y_i|\mathbf{X}_i) = \bar{\beta}_0(\tau) + \bar{\beta}_1(\tau)X_{i1} + \bar{\beta}_2(\tau)X_{i2} + \dots + \bar{\beta}_k(\tau)X_{ik} = \bar{\beta}(\tau)\mathbf{X}_i$$

y cumple lo siguiente:

$$\bar{\beta}(\tau) = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \phi_\tau(Y_i - e_\tau(Y_i|\mathbf{X}_i))$$

Proposición 2. Dado ES_α el Expected Shortfall de una variable aleatoria X de grado α , podemos expresarlo en función de e_τ el expectil ($ES_\alpha = g(e_\tau)$) con e_τ cumpliendo $P(X \leq e_\tau) = \alpha$.

Demostración. En primer lugar definimos ambos conceptos. Vamos a tomar X la variable aleatoria que representa unas ganancias. El Expected Shortfall (ES) es la esperanza de los beneficios que no superan el VaR, es decir:

$$ES_\alpha = \mathbb{E}(X|X \leq VaR_\alpha) = \mathbb{E}(X|X \leq F^{-1}(\alpha))$$

Por otro lado, el expectil es el valor que minimiza la función:

$$\begin{cases} \tau(X - e_\tau)^2 & \text{si } X \geq e_\tau \\ (1 - \tau)(X - e_\tau)^2 & \text{si } X < e_\tau \end{cases}$$

Podemos definir también esta función como:

$$\begin{aligned} \phi_\tau(X - e_\tau) &= \tau(X - e_\tau)^2 I(X \geq e_\tau) + (1 - \tau)(X - e_\tau)^2 I(X < e_\tau) \\ &= (\tau I(X \geq e_\tau) + (1 - \tau)I(X < e_\tau))(X - e_\tau)^2 \end{aligned}$$

Vamos a desarrollar y obtener propiedades del expectil. Para ello, vamos a tomar θ y a ver cual es el valor de θ que minimiza la esperanza de la función.

$$g_{\tau,\theta}(X) = (\tau I(X \geq \theta) + (1 - \tau)I(X < \theta))(X - \theta)^2$$

Vemos su esperanza:

$$\begin{aligned} E_X(g_{\tau,\theta}(X)) &= E_X((\tau I(X \geq \theta) + (1 - \tau)I(X < \theta))(X - \theta)^2) = \\ &= \int_{-\infty}^{\infty} (\tau I(x \geq \theta) + (1 - \tau)I(x < \theta))(x - \theta)^2 f_X(x) dx = \\ &= \int_{-\infty}^{\theta} ((1 - \tau)(x - \theta)^2 f_X(x) dx + \int_{\theta}^{\infty} (\tau(x - \theta)^2 f_X(x) dx = \\ &= (1 - \tau) \int_{-\infty}^{\theta} (x - \theta)^2 f_X(x) dx + \tau \int_{\theta}^{\infty} (x - \theta)^2 f_X(x) dx \end{aligned}$$

Derivamos con respecto a theta

$$\begin{aligned} \frac{d}{d\theta} E_X(g_{\tau,\theta}(X)) &= (1 - \tau) \int_{-\infty}^{\theta} 2(x - \theta) f_X(x) dx + (1 - \tau)(\theta - \theta)^2 f_X(\theta) + \\ &\quad + \tau \int_{\theta}^{\infty} 2(x - \theta) f_X(x) dx + \tau(\theta - \theta)^2 f_X(\theta) = \\ &= (1 - \tau) \int_{-\infty}^{\theta} 2(x - \theta) f_X(x) dx + 2\tau \int_{\theta}^{\infty} (x - \theta) f_X(x) dx = \\ &= 2((1 - \tau)E_X((X - \theta) \cdot 1_{X \leq \theta}) + \tau E_X((X - \theta) \cdot 1_{X > \theta})) = 0 \end{aligned}$$

Por tanto, simplificando, tenemos que:

$$E_X((X - \theta) \cdot 1_{X \geq \theta}) = -\frac{\tau}{1 - \tau} E_X((X - \theta) \cdot 1_{X < \theta})$$

y sustituyendo θ por e_{τ} , tenemos que:

$$E_X((X - e_{\tau}) \cdot 1_{X \geq e_{\tau}}) = -\frac{1 - \tau}{\tau} E_X((X - e_{\tau}) \cdot 1_{X < e_{\tau}})$$

denotamos $A = E_X((X - e_{\tau}) \cdot 1_{X \geq e_{\tau}})$ y $B = E_X((X - e_{\tau}) \cdot 1_{X < e_{\tau}})$, por lo que:

$$A = -\frac{1 - \tau}{\tau} B$$

Sabiendo que $\alpha = P(X \leq e_{\tau})$, por lo que:

$$\begin{aligned} E_X((X - e_{\tau}) \cdot 1_{X \leq e_{\tau}}) &= \int_{-\infty}^{e_{\tau}} (x - e_{\tau}) f_X(x) dx = \int_{-\infty}^{e_{\tau}} x f_X(x) dx - e_{\tau} \int_{-\infty}^{e_{\tau}} f_X(x) dx = \\ &= \frac{\alpha}{\alpha} \int_{-\infty}^{e_{\tau}} x f_X(x) dx - e_{\tau} \alpha = \alpha(ES_{\alpha} - e_{\tau}) \end{aligned}$$

Por tanto, despejando ES_{α} :

$$ES_{\alpha} = \frac{1}{\alpha} E_X((X - e_{\tau}) \cdot 1_{X \leq e_{\tau}}) + e_{\tau}$$

Queremos una expresión más sencilla. Sumamos A y B, calculados anteriormente:

$$\begin{aligned} A + B &= E_X((X - e_{\tau}) \cdot 1_{X \geq e_{\tau}}) + E_X((X - e_{\tau}) \cdot 1_{X < e_{\tau}}) = \\ &= E_X(X - e_{\tau}) = E_X(X) - e_{\tau} \end{aligned}$$

Al mismo tiempo,

$$A + B = -\frac{1-\tau}{\tau}B + B = \left(1 - \frac{1-\tau}{\tau}\right)B = \frac{2\tau-1}{\tau}B$$

Por tanto, sustituyendo en la expresión anterior:

$$\begin{aligned} ES_\alpha &= \frac{1}{\alpha}B + e_\tau = \frac{1}{\alpha}E_X((X - e_\tau) \cdot 1_{X \leq e_\tau}) + e_\tau = \\ &= \frac{\tau}{\alpha \cdot (2\tau - 1)}(E_X(X) - e_\tau) + e_\tau \end{aligned}$$

□

Por tanto, utilizaremos la regresión expectil para estimar que valor de τ es el más adecuado para que el expectil e_τ coincida con el VaR ($P(X \leq e_\tau) = \alpha \implies e_\tau = \text{Var}_\alpha$). Una vez tengamos el expectil, podremos estimar el Expected Shortfall (ES).

2.2. Modelos basados en Machine Learning

Conforme al avance de la tecnología, se han ido desarrollando nuevos métodos de estimación del riesgo, algunos basados en el aprendizaje máquina, que se ha desarrollado en los últimos años exponencialmente. Los métodos basados en Machine Learning (ML) son modelos de muchas dimensiones que se combinan con métodos de “regularización” y de reducción del sobreajuste, junto con algoritmos eficientes de búsqueda.

La base de los modelos de ML es parecida a la de los modelos econométricos, ya que están basados en métodos estadísticos. Sin embargo existen diferencias entre los modelos que los vuelven muy diferentes.

Mientras que los modelos econométricos se centran en la inferencia causal y la interpretación de relaciones económicas, los modelos de ML se enfocan en la predicción y precisión del ajuste. Además, los modelos de ML no se basan en supuestos como la normalidad o la independencia, al contrario que muchos métodos econométricos, lo que los hace más fielmente aplicables a distintas situaciones.

Previo a la utilización de estos métodos, es necesario “entrenarlos”, es decir, ajustar los parámetros del modelo a los datos. Para ello, se dividen los datos en dos conjuntos: el conjunto de entrenamiento y el conjunto de prueba. El conjunto de entrenamiento se utiliza para ajustar el modelo, mientras que el conjunto de prueba se utiliza para evaluar su rendimiento. Un vez entrenamos el modelo, podremos aplicarlo sobre otros datos, obteniendo predicciones sobre el riesgo asociado a los eventos.

Se van a presentar a continuación modelos utilizados para la estimación de cuantiles y expectiles, que posteriormente utilizaremos para estimar las variables económicas anteriormente descritas.

2.2.1. Redes Neuronales

Las redes neuronales funcionan como cerebros humanos, en los que encontramos neuronas conectadas entre sí. Están formadas por varios conjuntos de neuronas, que llamamos capas. Hay tres tipos de capas, llamadas la capas de entrada, ocultas y de salida. Para entrenar a la red neuronal, se introducen los datos por la capa de entrada. A partir de ahí, la información se transporta por las capas ocultas mediante funciones predeterminadas y pesos que determinan la influencia entre neuronas interconectadas. Este proceso se repite hasta que la información llega a la capa de salida. El proceso de razonamiento interno

funciona como una caja negra y no podemos identificar realmente las relaciones entre la entrada y la salida.

Para hacer una explicación completa pero sencilla, tomaremos una red neuronal con una sola capa oculta totalmente conectada. Las entradas x_i $i = 1, 2, \dots, n$ son transformadas mediante pesos y una función llamada función de activación y transportadas a la siguiente neurona, la cual está en la capa oculta. Acto seguido, se repetirá el proceso entre la capa oculta y la capa de salida pero utilizando Z_j $j = 1, 2, \dots, M_1$ como entradas.

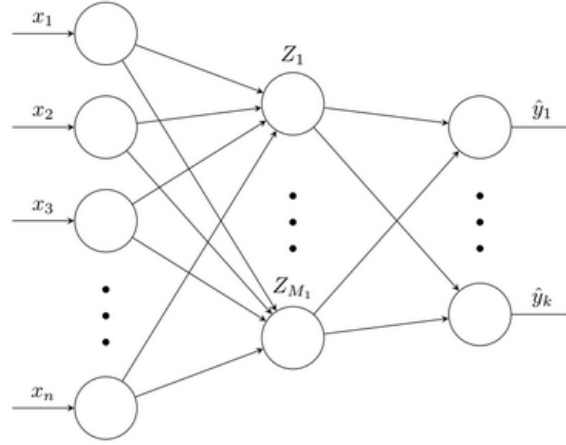


Figura 2.1: Red neuronal

Para cada neurona de la capa oculta, transformamos todos los valores de las neuronas de la capa de entrada de la siguiente manera:

$$a_j^{(1)} = \sum_{i=1}^n w_{j,i}^{(1)} x_i + b_j^{(1)}$$

$$Z_j^{(1)} = \sigma(a_j^{(1)}), j = 1, 2, \dots, M_1$$

con σ la función de activación y $w_{j,i}$ los pesos asociados a cada neurona, además de un peso de la propia neurona receptora, llamada sesgo b_j . En caso de haber más capas ocultas repetiríamos este proceso:

$$Z_j^{(m)} = \sigma \left(\sum_{i=1}^{M_{(m-1)}} w_{j,i}^{(m)} Z_i^{(m-1)} + b_j^{(m)} \right)$$

En este el caso de la figura 2.1, al solo tener una capa oculta, la salida de la red neuronal será:

$$\hat{y}_k = \sigma \left(\sum_{i=1}^{M_1} w_{k,i}^{(2)} Z_i^{(1)} + b_k^{(2)} \right)$$

Tras obtener los resultados de la primera evaluación, se actualizan los pesos de la red intentando minimizar la siguiente función:

$$\mathcal{L}(y, \hat{y}_i) = \sum_{i=1}^n \rho(y_i - \hat{y}_i)$$

Siendo y_i el valor esperado. Es decir, intentamos modificar los pesos de manera que la predicción sea lo más parecida al valor real posible. Modificaremos la función ρ en función de cual sea el objetivo de cálculo de la red.

En cuanto a las funciones de activación, la más utilizada e implementada en la mayoría de métodos es la función ReLu:

$$\sigma(x) = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} = \text{máx}(0, x)$$

Para ajustar los pesos, utilizaremos el algoritmo de retropropagación (*backpropagation*), que se basa en el cálculo del gradiente de la función de pérdida con respecto a los pesos. Sin embargo, en lugar de utilizar el método del gradiente descendente clásico, cuyo uso es sencillo y fácil de comprender, vamos a emplear el algoritmo *Adam* (Adaptive Moment Estimation), que utiliza estimaciones adaptativas de la media y la varianza del gradiente para actualizar los parámetros de manera más eficiente.

Dado el gradiente de la función de pérdida con respecto a los pesos de la capa l , que denotamos por $g^{(l)} = \frac{\partial \mathcal{L}}{\partial \theta} = -\sum_{i=1}^n \rho'(y_i - \hat{y}_i) \cdot \frac{\partial \hat{y}_i}{\partial \theta}$, Adam calcula una estimación del primer momento (media) y del segundo momento (varianza) del gradiente mediante las siguientes ecuaciones:

$$\begin{aligned} m^{(l)} &\leftarrow \beta_1 m^{(l)} + (1 - \beta_1) g^{(l)} \\ v^{(l)} &\leftarrow \beta_2 v^{(l)} + (1 - \beta_2) (g^{(l)})^2 \\ m_b^{(l)} &\leftarrow \beta_1 m_b^{(l)} + (1 - \beta_1) g_b^{(l)} \\ v_b^{(l)} &\leftarrow \beta_2 v_b^{(l)} + (1 - \beta_2) (g_b^{(l)})^2 \end{aligned}$$

A continuación, se realiza una corrección del sesgo de estas estimaciones, importante en las primeras iteraciones:

$$\begin{aligned} \hat{m}^{(l)} &= \frac{m^{(l)}}{1 - (\beta_1)^t} & \hat{v}^{(l)} &= \frac{v^{(l)}}{1 - (\beta_2)^t} \\ \hat{m}_b^{(l)} &= \frac{m_b^{(l)}}{1 - (\beta_1)^t} & \hat{v}_b^{(l)} &= \frac{v_b^{(l)}}{1 - (\beta_2)^t} \end{aligned}$$

Finalmente, se actualizan los pesos y el sesgo

$$\begin{aligned} w^{(l)} &\leftarrow w^{(l)} - \zeta \cdot \frac{\hat{m}^{(l)}}{\sqrt{\hat{v}^{(l)} + \epsilon}} \\ b^{(l)} &\leftarrow b^{(l)} - \zeta \cdot \frac{\hat{m}_b^{(l)}}{\sqrt{\hat{v}_b^{(l)} + \epsilon}} \end{aligned}$$

donde ζ es la tasa de aprendizaje, β_1 y β_2 son los coeficientes de decaimiento para los momentos (típicamente 0.9 y 0.999), ϵ es un pequeño valor positivo para evitar divisiones entre 0 y t es la iteración actual.

Suponemos, por tanto, que tenemos m datos disponibles $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_m, y_m)$, donde \mathbf{x}_i son las variables explicativas y y_i la variable dependiente. El algoritmo será:

Algorithm 1: Red neuronal (Estructura Correcta)

```

// Inicialización;
1 Inicializar aleatoriamente los pesos  $w^{(l)}$ , sesgos  $b^{(l)}$ , y momentos  $m^{(l)}, v^{(l)}, m_b^{(l)}, v_b^{(l)}$  para
    $l = 1, \dots, L$ ;
2  $t \leftarrow 0$ ;
   // Bucle de entrenamiento;
3 for  $epoch = 1$  to  $T$  do
4   for  $i = 1$  to  $m$  do
5      $t \leftarrow t + 1$ ;
     ; // 1. Propagación hacia adelante
6      $Z^{(0)} \leftarrow \mathbf{x}_i$ ;
7     for  $l = 1$  to  $L$  do
8        $a^{(l)} \leftarrow w^{(l)}Z^{(l-1)} + b^{(l)}$ ;
9        $Z^{(l)} \leftarrow \sigma^{(l)}(a^{(l)})$ ;
10     $\hat{y}_i \leftarrow Z^{(L)}$ ;
     ; // 2. Retropropagación (Cálculo de gradientes)
11     $\delta^{(L)} \leftarrow \frac{\partial \mathcal{L}}{\partial \hat{y}_i} \cdot \sigma^{(L)'}(a^{(L)})$ ;
12     $g^{(L)} \leftarrow \delta^{(L)}(Z^{(L-1)})^T$ ;  $g_b^{(L)} \leftarrow \delta^{(L)}$ ;
13    for  $l = L - 1$  to  $1$  do
14       $\delta^{(l)} \leftarrow ((w^{(l+1)})^T \delta^{(l+1)}) \circ \sigma^{(l)'}(a^{(l)})$ ;
15       $g^{(l)} \leftarrow \delta^{(l)}(Z^{(l-1)})^T$ ;  $g_b^{(l)} \leftarrow \delta^{(l)}$ ;
     ; // 3. Actualización de parámetros (Optimizador Adam)
16    for  $l = 1$  to  $L$  do
       ; // Actualizar momentos
17       $m^{(l)} \leftarrow \beta_1 m^{(l)} + (1 - \beta_1)g^{(l)}$ ;  $m_b^{(l)} \leftarrow \beta_1 m_b^{(l)} + (1 - \beta_1)g_b^{(l)}$ ;
18       $v^{(l)} \leftarrow \beta_2 v^{(l)} + (1 - \beta_2)(g^{(l)})^2$ ;  $v_b^{(l)} \leftarrow \beta_2 v_b^{(l)} + (1 - \beta_2)(g_b^{(l)})^2$ ;
       ; // Corregir sesgos de los momentos
19       $\hat{m}^{(l)} \leftarrow \frac{m^{(l)}}{1 - \beta_1^t}$ ;  $\hat{m}_b^{(l)} \leftarrow \frac{m_b^{(l)}}{1 - \beta_1^t}$ ;
20       $\hat{v}^{(l)} \leftarrow \frac{v^{(l)}}{1 - \beta_2^t}$ ;  $\hat{v}_b^{(l)} \leftarrow \frac{v_b^{(l)}}{1 - \beta_2^t}$ ;
       ; // Actualizar parámetros
21       $w^{(l)} \leftarrow w^{(l)} - \zeta \frac{\hat{m}^{(l)}}{\sqrt{\hat{v}^{(l)} + \epsilon}}$ ;
22       $b^{(l)} \leftarrow b^{(l)} - \zeta \frac{\hat{m}_b^{(l)}}{\sqrt{\hat{v}_b^{(l)} + \epsilon}}$ ;

```

Una de sus principales ventajas es su simplicidad computacional y su capacidad para evitar el problema del gradiente desvanecido en regiones positivas, lo que mejora significativamente la eficiencia del entrenamiento mediante el método de descenso por gradiente. Además, su derivada es muy simple, lo que permite una propagación del gradiente más estable en comparación con funciones como la sigmoide.

Las redes neuronales tienen una gran capacidad de aprendizaje y alta precisión en tareas complejas. Estas son las características que hacen que su utilización sea muy extendida en modelos predictivos.

2.2.2. Algoritmos de Boosting

El boosting asume la existencia de algoritmos débiles que, dados ejemplos de entrenamiento etiquetados, produce un predictores débiles. Supone que a continuación, la combinación de estos predictores débiles producen un predictor fuerte de alta precisión.

El objetivo es ir generando estos predictores de forma iterativa. Para que sean diferentes, tendremos que modificar el peso o importancia de cada dato dentro de la muestra, para que cada clasificador débil se "especialice" en unos pocos datos. El peso de cada dato se modificará según la predicción hasta ese momento este cerca del valor real o no.

Los algoritmos de Boosting se basan en minimizar una función de pérdida de forma iterativa, ajustando cada nuevo modelo para corregir los errores de los modelos anteriores.

Para ajustar el modelo, debemos entrenarlo, al igual que ocurre con las redes neuronales. Por tanto, para la explicación del modelo, tomamos $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$, siendo \mathbf{x}_i variables explicativas y y_i la variable dependiente.

Al igual que anteriormente, el objetivo del algoritmo será minimizar una función de pérdida y, una vez obtenido el predictor débil que minimiza este error, lo añadiremos al predictor fuerte y modificaremos los pesos de los datos de entrenamiento. Denotaremos la función de pérdida de la siguiente manera:

$$L(y, f(x)) = \sum_{i=1}^n \rho(y_i - f(x_i))$$

La función de pérdida ρ será modificada en función del objetivo de cálculo del algoritmo de boosting. Para calcular cuantiles, utilizaremos la función de pérdida cuantil o *pinball loss* y para calcular expectiles utilizaremos la función de pérdida expectil, ambas ya presentadas anteriormente.

El algoritmo de boosting es ideal para afrontar el problema ya que \mathbf{x}_i es una variable multidimensional y el problema puede tornarse complejo. La eficiencia y precisión de los algoritmos de boosting los hacen muy útiles en la estimación de cuantiles y expectiles.

Vamos a utilizar árboles de regresión como predictores débiles. Un árbol de regresión dividirá el conjunto de datos de entrenamiento según los valores de las variables independientes. En cada nodo, el algoritmo elige la variable y el umbral que minimizan la pérdida cuantil o expectil.

Para la optimización de los predictores utilizaremos el gradiente descendente, explicado anteriormente. En este contexto, aplicamos el gradiente descendente para mejorar la función de predicción en cada iteración del algoritmo de boosting.

Además, deberemos dar un valor a lo que nuestro predictor final "aprende" en cada iteración, asignando un valor a lo que llamaremos la tasa de aprendizaje, η . En principio, el valor de la tasa de aprendizaje será un valor pequeño para que el predictor vaya aprendiendo poco a poco. Lo haremos de la siguiente manera:

Algorithm 2: Boosting para cuantiles y expectiles

```

1 Inicializar  $f_0(x) \leftarrow 0$  o  $f_0(x) \leftarrow \bar{y}$  (la media de los valores  $y_i$ );
2 for  $m = 1$  to  $M$  do
3   for  $i = 1$  to  $n$  do
4      $u_{i,m} \leftarrow -\rho'_\tau(y_i - f_{m-1}(x_i));$ 
5      $u_{i,m} \leftarrow \mathbb{I}(y_i - f_{m-1}(x_i) < 0)(\tau - 1);$ 
6   ; // Ajustar un predictor débil  $g_m(x)$  a los pares  $(x_i, u_{i,m})$ 
7    $\{(x_i, u_{i,m})\}_{i=1}^n \longrightarrow g_m(x);$ 
8    $f_m(x) \leftarrow f_{m-1}(x) + \eta \cdot g_m(x);$ 

```

Capítulo 3

Aplicaciones de modelos a datos de créditos

—————Todo esto ponerlo en la zona de aplicación para explicar los métodos—————

- Como ya hemos comentado, el modelo tiene asociados unos parámetros entre los que se encuentran la tasa de aprendizaje y la profundidad máxima que puede alcanzar un predictor débil (árbol de regresión). Estos parámetros son los que van a determinar la complejidad del modelo y su capacidad de ajuste. Para obtener los parámetros óptimos, utilizaremos una función de medida de la corrección del ajuste e iteraremos distintos valores de los parámetros.

La función de medida que vamos a utilizar es la función de pérdida, llamada "Mean Pinball Loss":

$$pinball(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(y_i - \hat{y}_i)$$

$$\hat{c}(x) = \begin{cases} -\sum_{i=0}^k b[i] \cdot x[i] - capacidad \cdot (p - k) & \text{si caben en el tren} \\ +\infty & \text{si no caben en el tren} \end{cases}$$

Para su elección, utilizaremos el método de validación cruzada. Este método consiste en dividir el conjunto de datos en k partes, de las cuales $k - 1$ se utilizan para entrenar el modelo y la restante para evaluar su rendimiento. Este proceso se repite k veces, cada vez utilizando una parte diferente como conjunto de prueba. Al final, se promedian los resultados obtenidos en cada iteración para obtener una estimación más robusta del rendimiento del modelo.

Capítulo 4

Medidas de riesgo

Incluimos en este capítulo algunos comentarios generales y consejos de escritura que no se refieren específicamente a esta plantilla.

4.1. VaR

Como en cualquier documento escrito en \LaTeX (en nuestro caso, *pdflatex*) podemos incluir figuras. Estas figuras pueden ser imágenes contenidas en documentos aparte o compuestas dentro de nuestro documento con órdenes de \LaTeX . Se pueden incluir imágenes en diversos formatos gráficos, como pdf, jpeg y png, pero no todos; por ejemplo, no se pueden incluir imágenes tiff o gif.

Es preferible dejar flotar las figuras. Si se usa el entorno *figure*, entonces con

```
\begin{figure}[ht]  
...  
\end{figure}
```

indicamos que la figura debe insertarse preferiblemente donde se escribe el código (la h de *here*) o en la parte superior de la página (la t de *top*).



Figura 4.1: dos imágenes en archivos externos.

El paquete *tikz* permite hacer gráficas de funciones, figuras geométricas y diagramas sencillos, y no tan sencillos, como los de las figuras 4.2 y 4.3.

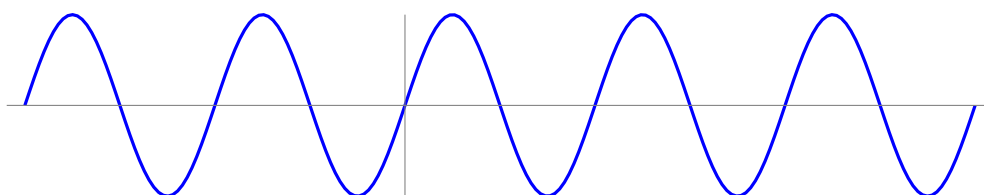
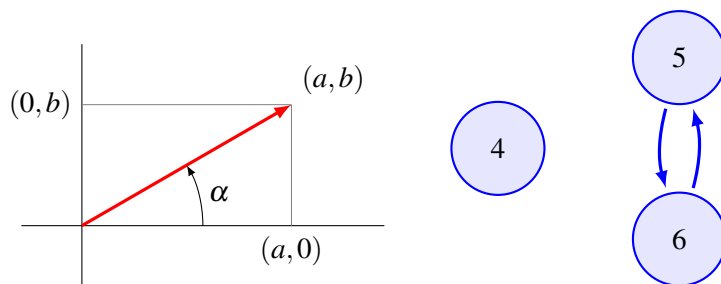


Figura 4.2: gráfica de la función $\sin(2x)$ entre -2π y 3π , hecha con *tikz*.

Figura 4.3: dos figuras juntas hechas con *tikz*.

4.2. CoVar

A menudo, los capítulos y secciones, los enunciados (teoremas, proposiciones, corolarios...) y las figuras llevan una numeración automática. Las fórmulas centradas pueden llevar también numeración o no llevarla. Es mejor numerar solo las fórmulas a las que se hace referencia en el texto. En cualquier caso, nunca se debe numerar *a mano*, sino usar las órdenes de \LaTeX adecuadas; por ejemplo,

```
\begin{equation}
...
\end{equation}
```

Y tampoco se debe hacer referencia a las fórmulas y enunciados escribiendo a mano la numeración, sino etiquetando la fórmula o enunciado con `\label` y refiriéndonos a ellos con `\eqref` (para fórmulas) y `\ref` (para enunciados).

De la misma manera, para hacer referencia a una entrada de la bibliografía debemos usar la orden `\cite`, nunca escribir a mano la clave.

4.3. Déficit esperado (Expected Shotfall)

Al escribir un trabajo hay algunas reglas de estilo a las que conviene atenerse. Es fácil encontrar textos, más o menos exhaustivos, que recogen estas reglas. Aunque en ocasiones no queramos seguirlas hasta el último detalle, es útil releerlas de vez en cuando.

En la bibliografía hemos seleccionado algunas referencias con indicaciones sobre \LaTeX , tipografía y lingüística ([3, 4, 12]).

Es singularmente interesante la página *texnia* (<http://www.texnia.com/>) de Javier Bezos, una autoridad en \LaTeX y tipografía. Ahí está disponible [4], entre mucha más información.

La guía [3] está escrita para los alumnos de la Universidad de La Rioja que redacten su trabajo de fin de grado en \LaTeX . De ella hemos tomado buena parte de estas notas. Uno de sus autores, Juan Luis Varona, es otro gran experto en \LaTeX , edición y tipografía.

Desde luego, para cualquier duda lingüística en castellano es indispensable la página de la Real Academia Española [12], donde encontraremos su diccionario de la lengua española y su diccionario panhispánico de dudas.

Recogemos aquí algunos primeros consejos:

1. Redacta correctamente. Aunque se trate de un texto matemático, debe estar bien escrito y sin faltas de ortografía o gramaticales.
2. Escribe un buen código \LaTeX . Un código claro, bien estructurado, es más fácil de repasar. Lo que parece sencillo de entender en el momento en que lo escribes, puede resultar oscuro si lo tienes que revisar al cabo de un tiempo. Pon comentarios (líneas que empiezan por `%`) para explicar las partes más complicadas.

3. Corrige todos los errores de \TeX que ocurran al componer el texto. No te conformes nunca con pulsar *Enter* hasta que salga el pdf. Intenta obtener al final una composición limpia, sin errores ni advertencias. En particular, sin avisos de *overflow*.
4. No incluyas una inmensa lista de paquetes o definiciones que no vas a usar, a menudo copiados y pegados de otros sitios.
5. Usa los entornos del tipo teorema, definición, observación... (es decir, con `\begin{teorema}`, ..., `\end{teorema}`), en lugar de formatear a mano los enunciados.
6. No cites nunca a mano las fórmulas y enunciados; no pongas numeración a fórmulas que no cites. Sobre esto hemos comentado algo en la sección 4.2.
7. Evita, si es posible, las órdenes `\bigskip`, `\medskip`, `\smallskip`, `\hspace`, `\vspace` y similares, y no uses `\`, `\`, y análogas sin una buena razón. Es decir, como regla general es mejor dejar que \LaTeX establezca el espaciado vertical y horizontal por sí mismo.
8. Haz un uso moderado de `\emph`, `\textbf` y similares: una palabra en cursiva o negrita llama la atención, pero una página llena de cursivas, negritas, texto entre comillas, recuadros o incluso distintos colores es simplemente más difícil de leer. Decide con qué criterio vas a usar cada uno de estos recursos tipográficos.
9. Consulta un buen manual, como [6] y [10], antes de escribir una expresión matemática nueva. Es probable que allí encuentres una solución más sencilla que la que vayas a desarrollar por tu cuenta. La mayoría de los paquetes tienen también su manual, que tendrás en tu distribución de \LaTeX .
10. Hay una gran variedad de órdenes para escribir fórmulas que no caben en una línea. Consulta la guía de usuario del paquete *amsmath*.
11. Establece un estilo para la bibliografía y sujétate a él. Esto también lo hemos comentado antes, en la sección ???. Ten en cuenta, para decidir el estilo, que las entradas de la bibliografía pueden ser de varios tipos: libros enteros, artículos de revistas, páginas web... Puedes tomar el estilo de alguna revista, por ejemplo.
12. No abrevies los nombres de las revistas de cualquier manera: hay formas normalizadas de hacerlo. Puedes consultar las abreviaturas de revistas matemáticas en [1], de la American Mathematical Society (AMS). También puedes consultar MathSciNet (por ejemplo, [2]), así mismo de la AMS y suscrita por la Universidad de Zaragoza. Si no encuentras la revista que quieres, puedes buscar en [9] las abreviaturas de palabras en títulos de revistas, en diferentes idiomas, según la norma ISO.

4.4. Modelos de riesgo

El contenido de la sección 4.4.1 está tomado, sin apenas cambios, de [5, pág. 518 y sig.]. Y el de la sección 4.4.3, de [7, pág. 185–187] y [13, pág. 206 y sig.]. De las muchas demostraciones de la fórmula de Stirling, la que presentamos en la sección ??? está tomada esencialmente de [11].

Para la primera sección es necesario tener conocimientos de análisis complejo. Damos por conocida la factorización

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right), \quad z \in \mathbb{C}, \quad (4.1)$$

cuya demostración puede verse, por ejemplo, en [5, pág. 518] o en [13, pág. 211–212].

Enunciamos aquí el teorema de factorización que usamos en la sección 4.4.1 (en realidad, el teorema de factorización de Weierstrass es más bien un recíproco de este resultado):

Definición. Escribimos $E_0(w) = 1 - w$ y para cada $p \in \mathbb{N}$,

$$E_p(w) = (1 - w) \exp \left(w + \frac{w^2}{2} + \frac{w^3}{3} + \cdots + \frac{w^p}{p} \right).$$

Estas funciones enteras se llaman *factores elementales* y su único cero es 1, simple.

Teorema 4.1. Sea $\{z_n\}_{n=1}^{\infty} \subseteq \mathbb{C}$ con $z_n \neq 0$ para todo n y $z_n \rightarrow \infty$ (pueden estar repetidos). Si $\{p_n\}_{n=1}^{\infty} \subseteq \mathbb{N} \cup \{0\}$ y para cada $r > 0$ la serie

$$\sum_{n=1}^{\infty} \left(\frac{r}{|z_n|} \right)^{1+p_n}$$

converge, entonces:

- a) $f(z) = \prod_{n=1}^{\infty} E_{p_n} \left(\frac{z}{z_n} \right)$ define una función holomorfa en \mathbb{C} ;
- b) los ceros de f son los z_n y su multiplicidad es el número de veces que aparecen repetidos en la sucesión.

4.4.1. Modelos clásicos

La función entera más sencilla que se anula, con ceros simples, en los enteros negativos es

$$F(z) = \prod_{n=1}^{\infty} \left(1 + \frac{z}{n} \right) e^{-z/n}. \quad (4.2)$$

4.4.2. Modelo GARCH

En efecto, basta aplicar el teorema de factorización con $z_n = -n$ para cada $n \in \mathbb{N}$ y con $p_n = 1$; claramente, para cada $r > 0$ la serie

$$\sum_{n=1}^{\infty} \left(\frac{r}{|z_n|} \right)^{1+p_n} = \sum_{n=1}^{\infty} \frac{r^2}{n^2}$$

converge, y entonces el producto infinito

$$\prod_{n=1}^{\infty} E_{p_n} \left(\frac{z}{z_n} \right) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{z_n} \right) e^{z/z_n} = \prod_{n=1}^{\infty} \left(1 + \frac{z}{n} \right) e^{-z/n}$$

converge uniformemente sobre compactos de \mathbb{C} y define una función entera cuyos ceros son los enteros negativos, todos ellos simples.

De la factorización (4.1) se deduce que

$$\operatorname{sen} \pi z = \pi z F(z) F(-z).$$

Las funciones $F(z-1)$ y $zF(z)$ son ambas enteras y con los mismos ceros: $0, -1, -2, \dots$, todos ellos simples. Por lo tanto, su cociente tiene logaritmo holomorfo: es una función entera que no se anula en ningún punto. Es decir:

$$F(z-1) = zF(z)e^{g(z)} \quad (4.3)$$

para alguna función entera g . Si sustituimos F por su definición, tenemos

$$\prod_{n=1}^{\infty} \left(1 + \frac{z-1}{n} \right) e^{-\frac{z-1}{n}} = ze^{g(z)} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n} \right) e^{-z/n}.$$

En esta expresión tomamos *derivadas logarítmicas*:

$$\sum_{n=1}^{\infty} \left(\frac{1/n}{1 + \frac{z-1}{n}} - \frac{1}{n} \right) = \frac{1}{z} + g'(z) + \sum_{n=1}^{\infty} \left(\frac{1/n}{1 + \frac{z}{n}} - \frac{1}{n} \right).$$

No insistiremos en esto, pero todos los productos infinitos, las series y los límites que hemos manejado y los que vengan a continuación convergen porque converge el producto infinito (4.2) y porque se puede derivar logarítmicamente.

De la fórmula anterior obtenemos $g'(z)$, pasando a sumas parciales para no manejar series que no converjan:

$$\begin{aligned} \frac{1}{z} + g'(z) &= \sum_{n=1}^{\infty} \left(\frac{1/n}{1 + \frac{z-1}{n}} - \frac{1/n}{1 + \frac{z}{n}} \right) = \lim_N \sum_{n=1}^N \left(\frac{1}{n+z-1} - \frac{1}{n+z} \right) \\ &= \lim_N \left(\frac{1}{z} - \frac{1}{N+z} \right) = \frac{1}{z}. \end{aligned}$$

Por consiguiente, $g'(z) = 0$ para todo $z \in \mathbb{C}$, es decir, g es una constante. ¿Qué constante? La llamamos γ y para hallarla basta imponer que cumpla la fórmula (4.3) sustituyendo, por ejemplo, $z = 1$:

$$1 = F(0) = F(1)e^{\gamma} = e^{\gamma} \prod_{n=1}^{\infty} \left(1 + \frac{1}{n} \right) e^{-1/n},$$

es decir,

$$e^{-\gamma} = \prod_{n=1}^{\infty} \left(1 + \frac{1}{n} \right) e^{-1/n}.$$

Naturalmente, no hay un único valor porque el logaritmo complejo no es único, pero el producto infinito de la derecha es un número real positivo y entonces podemos elegir el único $\gamma \in \mathbb{R}$ que cumple esa relación. Es decir:

$$\gamma = -\log \prod_{n=1}^{\infty} \left(1 + \frac{1}{n} \right) e^{-1/n} = -\log \lim_N \prod_{n=1}^N \frac{n+1}{n} e^{-1/n} = -\log \lim_N (N+1) e^{-1-\frac{1}{2}-\dots-\frac{1}{N}}.$$

O, equivalentemente:

$$\gamma = \lim_N \left(1 + \frac{1}{2} + \dots + \frac{1}{N} - \log(N+1) \right).$$

Esta constante se llama *constante de Euler*. Observemos que como $\log(N+1) = \log N + \log(1 + \frac{1}{N})$, también podemos escribir

$$\gamma = \lim_N \left(1 + \frac{1}{2} + \dots + \frac{1}{N} - \log N \right). \quad (4.4)$$

Volviendo a la fórmula (4.3), pero ya con $g(z) = \gamma$, tenemos

$$F(z-1) = zF(z)e^{\gamma}$$

y por lo tanto

$$e^{\gamma(z-1)} F(z-1) = ze^{\gamma z} F(z).$$

La función $\Gamma(z) = \frac{1}{ze^{\gamma z} F(z)}$, cuyos polos son $0, -1, -2, \dots$, todos ellos simples, cumple

$$\begin{aligned} \Gamma(z+1) &= \frac{1}{e^{\gamma z} F(z)} = z\Gamma(z), \\ \Gamma(1) &= \frac{1}{e^{\gamma} F(1)} = 1. \end{aligned}$$

En particular, de aquí se deduce por inducción que $\Gamma(n) = (n-1)!$ para todo $n \in \mathbb{N}$. También se cumple que

$$\operatorname{sen} \pi z = \pi z F(z) F(-z) = \pi z \frac{1}{ze^{\gamma z} \Gamma(z)} \cdot \frac{1}{-ze^{-\gamma z} \Gamma(-z)} = \frac{\pi}{\Gamma(z)(-z)\Gamma(-z)} = \frac{\pi}{\Gamma(z)\Gamma(1-z)},$$

es decir,

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\operatorname{sen} \pi z}, \quad z \notin \mathbb{Z}.$$

Si hacemos $z = 1/2$, obtenemos $\Gamma(1/2)^2 = \pi$. Como

$$\Gamma\left(\frac{1}{2}\right) = \frac{1}{\frac{1}{2}e^{\frac{1}{2}\gamma}F\left(\frac{1}{2}\right)} = \frac{1}{\frac{1}{2}e^{\frac{1}{2}\gamma}\prod_{n=1}^{\infty}\left(1+\frac{1}{2n}\right)e^{-\frac{1}{2n}}} > 0,$$

se deduce que

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Por otra parte, de la definición de $\Gamma(z)$ se sigue que

$$\Gamma(z) = \frac{e^{-\gamma z}}{z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{z/n}, \quad z \neq 0, -1, -2, \dots$$

Por lo tanto,

$$\begin{aligned} \Gamma(z) &= \frac{e^{-\gamma z}}{z} \lim_N \prod_{n=1}^N \left(1 + \frac{z}{n}\right)^{-1} e^{z/n} = \lim_N \frac{e^{-\gamma z N!} e^{z(1+\frac{1}{2}+\dots+\frac{1}{N})}}{z(z+1)\dots(z+N)} \\ &= \lim_N \frac{e^{z \log N} N!}{z(z+1)\dots(z+N)} e^{z(-\gamma+1+\frac{1}{2}+\dots+\frac{1}{N}-\log N)}. \end{aligned}$$

Teniendo en cuenta (4.4), queda

$$\Gamma(z) = \lim_N \frac{N^z N!}{z(z+1)\dots(z+N)}, \quad z \neq 0, -1, -2, \dots$$

La función Γ se llama *función Γ de Euler*.

4.4.3. Modelos de aprendizaje automático

La definición más clásica de la función Γ es como una integral:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, \quad \operatorname{Re} z > 0.$$

Vamos a ver que esta definición coincide con la anterior. De momento, definamos

$$G(z) = \int_0^{\infty} t^{z-1} e^{-t} dt.$$

4.4.4. Modelo XGBoost

Como $|t^{z-1} e^{-t}| = t^{\operatorname{Re}(z)-1} e^{-t}$, se deduce que la función G está bien definida si $\operatorname{Re} z > 0$. En realidad, es fácil comprobar que es una función holomorfa en el abierto $\{z \in \mathbb{C}; \operatorname{Re} z > 0\}$. Algunas propiedades sencillas de esta función son:

- a) $G(z+1) = zG(z)$ (se comprueba integrando por partes).
- b) $G(1) = 1$ (es una integral inmediata).

- c) Para cada $n \in \mathbb{N}$, $G(n) = (n-1)!$ (se prueba por inducción, a partir de las dos propiedades anteriores).

Sean ahora $x, y > 0$ (números reales positivos), $0 < \lambda < 1$. Entonces,

$$G(\lambda x + (1-\lambda)y) = \int_0^\infty t^{\lambda x + (1-\lambda)y-1} e^{-t} dt = \int_0^\infty t^{\lambda(x-1)} e^{-\lambda t} t^{(1-\lambda)(y-1)} e^{-(1-\lambda)t} dt.$$

Aplicamos ahora la desigualdad de Hölder a las funciones $t^{\lambda(x-1)} e^{-\lambda t}$, $t^{(1-\lambda)(y-1)} e^{-(1-\lambda)t}$, con exponentes $p = 1/\lambda$, $q = 1/(1-\lambda)$. Resulta:

$$\begin{aligned} G(\lambda x + (1-\lambda)y) &\leq \left(\int_0^\infty t^{\lambda p(x-1)} e^{-\lambda p t} dt \right)^{1/p} \left(\int_0^\infty t^{(1-\lambda)q(y-1)} e^{-(1-\lambda)q t} dt \right)^{1/q} \\ &= \left(\int_0^\infty t^{x-1} e^{-t} dt \right)^\lambda \left(\int_0^\infty t^{y-1} e^{-t} dt \right)^{1-\lambda}. \end{aligned}$$

Es decir:

$$G(\lambda x + (1-\lambda)y) \leq G(x)^\lambda G(y)^{1-\lambda}, \quad x > 0, y > 0, 0 \leq \lambda \leq 1$$

(los casos $\lambda = 0$, $\lambda = 1$ son triviales). Las funciones que cumplen esta propiedad se llaman *logarítmicamente convexas*.

La demostración que vamos a ver de que G es la función Γ se aplica a cualquier función $G : (0, \infty) \rightarrow (0, \infty)$ que cumpla las propiedades *a*), *b*), *c*) y la convexidad logarítmica; es el teorema de Bohr-Mollerup. En nuestro caso, como G y Γ son holomorfas, si coinciden en $(0, \infty)$ coinciden en todo el abierto conexo $\{z \in \mathbb{C}; \operatorname{Re} z > 0\}$, por el principio de prolongación analítica.

Comencemos por un $x \in (0, 1]$. Sea $n \in \mathbb{N}$. Entonces,

$$n+1+x = (1-x)(n+1) + x(n+2),$$

luego, por la convexidad logarítmica (por eso hace falta $0 < x \leq 1$),

$$\begin{aligned} G(n+1+x) &\leq G(n+1)^{1-x} G(n+2)^x = G(n+1)^{1-x} [(n+1)G(n+1)]^x \\ &= (n+1)^x G(n+1) = (n+1)^x n! \end{aligned}$$

De la misma manera,

$$n+1 = x(n+x) + (1-x)(n+1+x),$$

luego

$$\begin{aligned} n! &= G(n+1) \leq G(n+x)^x G(n+1+x)^{1-x} \\ &= \left[\frac{G(n+1+x)}{n+x} \right]^x G(n+1+x)^{1-x} = \frac{G(n+1+x)}{(n+x)^x}. \end{aligned}$$

Juntando las dos desigualdades tenemos

$$(n+x)^x n! \leq G(n+1+x) \leq (n+1)^x n!,$$

es decir:

$$\left(1 + \frac{x}{n}\right)^x \leq \frac{G(n+1+x)}{n! n^x} \leq \left(1 + \frac{1}{n}\right)^x.$$

Si hacemos $n \rightarrow \infty$, deducimos que

$$1 = \lim_n \frac{G(n+1+x)}{n! n^x} = \lim_n \frac{x(x+1)(x+2) \dots (x+n)G(x)}{n! n^x},$$

donde en el último paso se usa la propiedad a). Por lo tanto,

$$G(x) = \lim_n \frac{n! n^x}{x(x+1)(x+2)\dots(x+n)} = \Gamma(x).$$

Ya tenemos que $G(x) = \Gamma(x)$ si $x \in (0, 1]$. Para $x > 1$ tendremos $x \in (m, m+1]$ para algún $m \in \mathbb{N}$ y entonces

$$G(x) = (x-1)(x-2)\dots(x-m)G(x-m) = (x-1)(x-2)\dots(x-m)\Gamma(x-m) = \Gamma(x).$$

Partimos de la definición de la función Γ como una integral, pero nos limitamos a valores reales:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x > 0.$$

Haciendo el cambio de variable $u = \sqrt{t}$, resulta

$$\Gamma(x) = \int_0^\infty 2u^{2x-1} e^{-u^2} du.$$

Por lo tanto,

$$\frac{\Gamma(x)}{x^{x-\frac{1}{2}} e^{-x}} = \int_0^\infty 2 \left(\frac{u}{\sqrt{x}} \right)^{2x-1} e^{x-u^2} du = \int_{-\sqrt{x}}^\infty 2 \left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} e^{-2v\sqrt{x}} e^{-v^2} dv,$$

donde para llegar a la segunda integral hacemos el cambio $u = \sqrt{x} + v$ (así que $x - u^2 = -2v\sqrt{x} - v^2$). Es decir,

$$\frac{\Gamma(x)}{x^{x-\frac{1}{2}} e^{-x}} = \int_{-\infty}^\infty \chi_{[-\sqrt{x}, +\infty)}(x) 2 \left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} e^{-2v\sqrt{x}} e^{-v^2} dv. \quad (4.5)$$

Queremos hacer $x \rightarrow +\infty$ usando el teorema de la convergencia dominada, para lo cual debemos encontrar una acotación y hallar un límite del integrando. En todo lo que sigue podemos suponer que $2x-1 > 0$. Empezamos con la acotación: usando la desigualdad

$$1+t \leq e^t, \quad \forall t \in \mathbb{R},$$

que se demuestra muy fácilmente, deducimos que

$$\left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} \leq e^{2v\sqrt{x} - \frac{v}{\sqrt{x}}}$$

y, por lo tanto,

$$0 \leq 2 \left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} e^{-2v\sqrt{x}} \leq 2e^{-v/\sqrt{x}} \leq 2e, \quad \text{si } v \in [-\sqrt{x}, +\infty).$$

Así que

$$0 \leq \chi_{[-\sqrt{x}, +\infty)}(x) 2 \left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} e^{-2v\sqrt{x}} e^{-v^2} \leq 2e e^{-v^2}.$$

Ya tenemos la acotación. Ahora, fijado $v \in [-\sqrt{x}, +\infty)$,

$$2 \left(1 + \frac{v}{\sqrt{x}} \right)^{2x-1} e^{-2v\sqrt{x}} = 2e^{(2x-1)\log(1+\frac{v}{\sqrt{x}}) - 2v\sqrt{x}},$$

y

$$\begin{aligned}\lim_{x \rightarrow +\infty} \left((2x-1) \log\left(1 + \frac{v}{\sqrt{x}}\right) - 2v\sqrt{x} \right) &= \lim_{x \rightarrow +\infty} \left(2x \log\left(1 + \frac{v}{\sqrt{x}}\right) - 2v\sqrt{x} \right) \\ &= \lim_{x \rightarrow +\infty} 2v^2 \frac{\log\left(1 + \frac{v}{\sqrt{x}}\right) - \frac{v}{\sqrt{x}}}{\frac{v^2}{x}} \\ &= \lim_{y \rightarrow 0^+} 2v^2 \frac{\log(1+y) - y}{y^2} = -v^2\end{aligned}$$

(por el desarrollo de Young de $\log(1+y)$ o por la regla de L'Hôpital). Es decir, para todo $v \in \mathbb{R}$,

$$\lim_{x \rightarrow +\infty} \chi_{[-\sqrt{x}, +\infty)}(x) 2 \left(1 + \frac{v}{\sqrt{x}}\right)^{2x-1} e^{-2v\sqrt{x}} e^{-v^2} = 2e^{-2v^2}.$$

Llevando todo a (4.5), obtenemos que

$$\lim_{x \rightarrow +\infty} \frac{\Gamma(x)}{x^{x-\frac{1}{2}} e^{-x}} = \int_{-\infty}^{+\infty} 2e^{-2v^2} dv.$$

Esta integral se calcula de varias maneras. Por ejemplo, usando coordenadas polares,

$$\begin{aligned}\left(\int_{-\infty}^{+\infty} 2e^{-2v^2} dv \right)^2 &= \iint_{\mathbb{R}^2} 4e^{-2(u^2+v^2)} du dv = \int_{(0,+\infty)} \int_{-\pi}^{\pi} 4e^{-2r^2} r d\phi dr \\ &= \int_{(0,+\infty)} 8\pi r e^{-2r^2} dr = \left[-2\pi e^{-2r^2} \right]_{r=0}^{r \rightarrow +\infty} = 2\pi,\end{aligned}$$

de donde

$$\int_{-\infty}^{+\infty} 2e^{-2v^2} dv = \sqrt{2\pi}.$$

Así que finalmente obtenemos:

$$\lim_{x \rightarrow +\infty} \frac{\Gamma(x)}{x^{x-\frac{1}{2}} e^{-x}} = \sqrt{2\pi},$$

que es la fórmula de Stirling. Otra manera equivalente de escribirla, más habitual, es esta:

$$\lim_{x \rightarrow +\infty} \frac{\Gamma(x+1)}{x^x e^{-x} \sqrt{2\pi x}} = \lim_{x \rightarrow +\infty} \frac{x\Gamma(x)}{x^{x+\frac{1}{2}} e^{-x} \sqrt{2\pi}} = \lim_{x \rightarrow +\infty} \frac{\Gamma(x)}{x^{x-\frac{1}{2}} e^{-x} \sqrt{2\pi}} = 1.$$

Capítulo 5

Aplicación de QR y XGBoost a ...

Capítulo 6

Análisis de resultados

Capítulo 7

Conclusión

Bibliografía

- [1] AMERICAN MATHEMATICAL SOCIETY, *Abbreviations of names of serials*, <http://www.ams.org/msnhtml/serials.pdf>.
- [2] AMERICAN MATHEMATICAL SOCIETY, *MathSciNet* (servidor de Estrasburgo), <http://ams.u-strasbg.fr/mathscinet/>.
- [3] A. ARENAS, E. LABARGA Y J. L. VARONA, *Una breve guía para redactar un trabajo de fin de grado con contenido matemático utilizando LaTeX*, https://www.unirioja.es/facultades_escuelas/fct/TFG/guia-TFG-matematicas.pdf, disponible en https://www.unirioja.es/facultades_escuelas/fct/TFG/guias_manuales.shtml.
- [4] J. BEZOS, *Ortotipografía y notaciones matemáticas*, <http://www.texnia.com/archive/ortomatem.pdf>, disponible en <http://www.texnia.com/>.
- [5] J. BRUNA Y J. CUFÍ, *Anàlisi complexa*, Colección Manuals, Universitat Autònoma de Barcelona, Bellaterra, 2008.
- [6] B. CASCALES, P. LUCAS, J. M. MIRA, A. PALLARÉS Y S. SÁNCHEZ-PEDREÑO, *LaTeX, una imprenta en tus manos*, Aula Documental de Investigación, Madrid, 2000.
- [7] B. CUARTERO, *Fundamentos de análisis matemático*, Colección Textos Docentes, Prensas Universitarias de Zaragoza, 2009.
- [8] FACULTAD DE CIENCIAS DE LA UNIVERSIDAD DE ZARAGOZA, *Directrices propias para la elaboración del trabajo fin de grado en Matemáticas*, disponible en <https://ciencias.unizar.es/trabajo-fin-de-grado-en-matematicas>.
- [9] INTERNATIONAL STANDARD SERIAL NUMBER, *International identifier for serials*, <http://www.issn.org/services/online-services/access-to-the-ltwa/>.
- [10] F. MITTELBAACH, M. GOOSSENS, J. BRAAMS, D. CARLISLE Y C. ROWLEY, *The L^AT_EX Companion*, 2.^a ed., Addison-Wesley, 2004.
- [11] J. M. PATIN, A very short proof of Stirling's formula, *Amer. Math. Monthly* **96** (1) (1989), 41–42.
- [12] REAL ACADEMIA ESPAÑOLA, <http://www.rae.es/>.
- [13] R. WEBSTER, *Convexity*, Oxford University Press, Oxford, 1994.