UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques i Informàtica

Treball Final de Grau

GRAU DE MATEMÀTIQUES

Risc de crèdit

Miquel Panadés Codina

Director: **Dr. Josep Vives i Santa-Eulàlia**
Facultat de Matemàtiques i Informàtica

Barcelona, Gener 2023

Resum

Les persones conviuen diàriament amb situacions on l'exposició al perill és present, i són les primeres que no volen que els hi passi res negatiu... Però pren tothom les mateixes precaucions? Està tothom exposat al mateix nivell risc?

La resposta a aquestes dues preguntes és òbviament no, i amb eines i models estadístics, l'objectiu és quantificar i preveure tant acuradament com sigui possible el risc que pateix una persona davant de diferents esdeveniments.

Aquest projecte es centra en un d'aquests riscs: el financer, i més concretament, en el risc de crèdit. Els bancs necessiten saber quin risc prenen quan concedeixen préstecs, quants diners han de reservar per a que el possible incompliment d'un client no els passi massa factura, quin interès hi han de posar...

Aquesta precaució és molt més present sobretot des de la crisi financera global del 2007-2008, i els bancs hi estan dedicant cada cop més recursos, cosa que beneficia i impulsa l'estudi de noves metodologies, sense deixar de fer servir els models tradicionals.

Abstract

People are constantly living exposed to danger, and they are the first ones that don't want any negative event happening to them... But does everyone have the same backstops? Is everyone exposed to the same amount of risk?

The answer to both these questions is obviously no, and with statistical tools and models, the goal is to quantify and predict as accurately as possible the risk that an individual is exposed to when facing different events.

This project focuses on one of these risks: the financial one, and more specifically, credit risk. Banks want to know what is the risk they are facing when granting loans, how much money they have to set aside so that the possible default of a customer does not cost them too much, the type of interest they have to put on the loan...

This caution is especially present since the global financial crisis of 2007-2008, and banks are devoting more and more resources to it, which benefits and drives the study of new methodologies, without ceasing to use tradicional models.

2020 Mathematics Subject Classification. 91G40, 91G70, 60G07

Agraïments

Primerament, m'agradaria agrair al meu tutor del treball, el Dr. Josep Vives i Santa-Eulàlia el seguiment tan adequat i les correccions i propostes brindades durant tot el procés del treball, que va iniciar-se fa un any.

A més, voldria agrair tant a la meva família, com als meus amics i a la meva companya de vida, tots els ànims, suport i bons moments que m'han donat durant aquests anys. M'encantaria que sentissin també aquesta sensació de felicitat i descans que sento jo d'haver acabat el grau de matemàtiques, ja que sense ells no hagués estat possible.

Gràcies, i espero seguir tenint-vos sempre.

Índex

Introducció

L'onze de setembre del 2001, tot just tocats tres quarts de nou del matí, l'avió del vol número 11 d'American Airlines s'estavellava contra la Torre Nord del World Trade Center. Era el primer de quatre avions segrestats per terroristes suïcides d'Al-Qaeda, que havien rebut fins i tot formació en pilotatge d'avions.

Les següents hores van ser caòtiques, els següents dies esfereïdors i els següents mesos molt dolorosos. Però tot i que sembli injust per les víctimes directes, quan realment aquesta tragèdia ha afectat més a nivell global ha estat en tots els anys des de llavors i fins a dia d'avui.

Començant per un malestar general dels americans, que es veien molt pessimistes i inestables, i que va provocar que des del 2001 fins al 2004 els bancs baixessin els interessos en els crèdits fins a rondar l'1%, amb l'objectiu de reactivar l'economia i alhora incentivar a les famílies i empreses a consumir i comprar més. Això la població americana ho va aprofitar sobretot per a la compra massiva d'habitatges, ja fos per necessitat o bé per especulació.

Mentre els prestataris poguessin tornar els crèdits i els habitatges anessin pujant de preu com fins llavors havia passat, no hi hauria cap problema i tot seguiria igual. Però òbviament, l'economia és un món on és molt difícil que hi guanyi tothom, i es va crear un excés de despeses i la inflació va pujar desmesuradament.

Un dels problemes principals va ser que els bancs atorgaven crèdits hipotecaris a clients sense el recolzament suficient, és a dir, amb nivells baixos d'ingressos i alt risc de morositat, anomenats persones amb baixa qualificació creditícia.

Davant d'aquest problema, els bancs van començar a pujar els interessos, des de finals de 2004 fins al 2006 van passar de menys de l'1% a més del 5%, i com que una gran part dels crèdits hipotecaris eren amb interès variable, els clients cada cop havien de pagar més, i com que eren persones amb baixa qualificació creditícia, molts d'ells van deixar de retornar els préstecs: les empreses es declaraven en fallida i els avaladors estaven en la mateixa situació de baixada econòmica.

Llavors els propietaris posaven els pisos a la venda, però amb els interessos dels bancs tan elevats, poca gent comprava i hi havia molta més oferta que demanda, així que els pisos van començar a baixar de valor.

A més, com que vivim en un món cada cop més globalitzat, la situació va estendre's ràpidament per tot el món, ja que molts bancs havien venut aquestes hipoteques a fons d'inversió internacionals en paquets, però com que tot es basava en que els prestataris tornessin els diners i els habitatges seguissin pujant de preu, i cap de les dues coses estava passant, la bombolla va explotar i la crisi del 2007-2008 va començar.

Conseqüentment, durant els següents anys hi va haver una gran quantitat d'empreses i particulars arreu del món que no van aguantar l'impacte econòmic de la crisi i van haver de declarar-se en fallida, deixant al descobert a les entitats financeres prestadores. És per això que cada cop s'inverteix més temps i diners en trobar nous algoritmes i noves tecnologies que optimitzin la presa de decisions a l'hora de concedir un préstec, quantificant el risc creditici per a cada individu o empresa i ajustant-ne tant l'interès per cada cas com les reserves de capital que han de procurar tenir, per a evitar una pèrdua de beneficis significativa.

Amb tota aquesta informació, el propòsit d'aquest treball és endinsar-se com a principiant en aquest món, solidificar conceptes econòmics al voltant dels crèdits i tot el que els envolta, i en especial en el risc que prenen les institucions prestadores quan els concedeixen. També es vol estudiar teoria de processos estocàstics, en particular conceptes de càlcul estocàstic, amb l'objectiu de caracteritzar els models de risc de crèdit tradicionalment més utilitzats, i comparar-los amb els potencialment millors: els models de *Machine Learning*.

Aquest projecte s'inicia amb una introducció al risc creditici, primer uns conceptes financers necessaris per a entendre el projecte, i després unes definicions i proposicions de modelització financera. Seguint amb la modelització, el segon és un capítol en el que s'introdueixen conceptes de càlcul estocàstic que seran ben útils per al següent capítol: models de risc de crèdit. En aquest es presenten diferents tipus de models financers tradicionals per a avaluar el risc de crèdit de qualsevol tipus de client, on hi apareixen conceptes matemàtics ja descrits en els capítols anteriors. Aquest capítol dóna pas a un altre de més modern: el *Machine Learning* en models d'aquest tipus. Aquí s'expliquen noves tècniques sobre aquest camp i quines són les principals maneres d'avaluar-ne els resultats. Finalment, s'expliquen les conclusions del treball i tot el que s'ha après amb l'experiència d'aprofundir en un tema que cada cop serà més important, i del que no en sabia res fa uns mesos.

Capítol 1

Introducció al Risc de Crèdit

En aquest capítol s'introdueixen alguns dels conceptes més essencials en els que es basa aquest treball i es responen les preguntes inicials per a entendre la matèria. Com per exemple: què és el risc?

Segons l'Enciclopèdia Catalana és *"la contingència desfavorable a la qual està exposat algú o alguna cosa; perill incert"*. En altres paraules, és l'exposició a una situació que té una certa probabilitat de perill, o en la que existeix la possibilitat de que ocorri un esdeveniment indesitjat per a l'individu exposat. Per tant, podria definir-se també com un valor quantitatiu per a mesurar aquesta possibilitat com a probabilitat i tractar-ne els temes relacionats com a esdeveniments aleatoris, i estudiar-los estadísticament.

Al tractar-se d'una mesura quantitativa i degut a l'existència de factors externs, anomenats factors de risc, aquest valor es pot reduir gairebé a zero afegint prevencions, per exemple per a evitar danys majors en accidents de trànsit existeix l'Airbag dels cotxes o bé per a evitar robatoris o per a detectar els lladres existeixen càmeres de vigilància en els habitatges. Existeixen moltes empreses dedicades a idear i fabricar aquestes prevencions, ja que les persones conviuen diàriament amb situacions on l'exposició al perill és present, i són les primeres que no volen que els hi passi res dolent... Però pren tothom les mateixes precaucions?

Activitats tan rutinàries com creuar un semàfor en vermell, no portar casc anant amb bicicleta o viure a prop d'una central nuclear ja són exposicions a esdeveniments desfavorables que en alguns casos es podrien evitar però en d'altres són inevitables, depenent de cada individu. Això ens fa veure que l'exposició al risc no és sempre la mateixa, i que potser un noi en plena joventut tendeix més a realitzar activitats perilloses que una senyora jubilada; o que una persona que viu al Japó en un arxipèlag propens a patir terratrèmols i tsunamis està més exposat que una persona que viu a un poble de Catalunya central; o que una persona amb algun tipus d'addicció tendeix a tenir més problemes amb els diners que una persona sana.

Amb tots aquests exemples no s'està donant una opinió personal, simplement s'estan posant sobre la taula fets que semblen evidents però que alhora porten molts mals de cap a les asseguradores, entre d'altres, i que sovint porten a debats ètics i morals. I és que aquests fets no deixen de ser, al cap i a la fi, propis de l'aleatorietat de la vida, que marca el destí de cadascú, el problema és que aquesta aleatorietat a vegades està esbiaixada.

És també evident que existeix un gran ventall de tipus de riscos, entre els quals es podrien trobar els biològics, físics i financers. En el món laboral tots són molt presents, i els sindicats de treballadors intenten assegurar que tots estiguem ben coberts davant dels dos primers tipus anomenats. Però en aquest projecte el que ens interessa més és precisament l'últim d'ells, el risc financer, amb el qual entendrem el risc de crèdit i les motivacions per a prevenir-lo tant com es pugui.

1.1 Risc financer

L'anàlisi del risc financer és un dels aspectes més importants per a totes les empreses, inversors, o qualsevol altra persona o grup que pugui patir algun esdeveniment amb conseqüències econòmiques, especialment si són negatives.

El risc financer fa referència a la incertesa en el rendiment d'una inversió deguda a canvis en els sectors on s'opera, a la impossibilitat de la devolució del capital per alguna de les parts involucrades o bé a causa de la inestabilitat del mercat. És la possibilitat de pèrdua monetària o de capital al què es podria arribar si les circumstàncies fossin desfavorables. És per això que és molt important per a tots els involucrats reduir aquesta exposició al risc, o bé tenir-la en compte i reservar recursos per a prevenir impactes greus.

Es distingeixen quatre tipus de riscos financers bàsics:

- Risc de mercat: És el risc que es pren si es compren actius financers en el mercat, donat que no es pot assegurar el retorn de la inversió. És el més comú i un dels més importants que existeixen ja que els seus efectes són molt amplis, degut a que es veu afectat directament per la relació entre l'oferta i la demanda del mercat, que varien constantment. Aquesta dinàmica respon, en gran part, a les incerteses econòmiques globals del moment, que afecten considerablement al rendiment de totes les empreses. Existeixen tres factors claus que son causes d'aquest tipus de risc:
  - (i) El tipus d'interès: Es tracta del risc associat als moviments en contra del tipus d'interès. En funció de la situació de l'empresa o de l'individu, pot interessar més o menys la pujada o baixada d'aquesta variable.
  - (ii) El tipus de canvi: Fa referència a les possibles variacions en la taxa de valor de les divises en el mercat.
  - (iii) Oferta i demanda: Afecta en gran part als accionistes i inversos, als que els hi disminueix el valor de la seva cartera si hi ha una davallada del valor de les accions que posseeixen.
- Risc de liquiditat: Aquest tipus de risc es refereix a la possibilitat de que una empresa sigui incapaç de complir les obligacions contractuals que té, perquè no té un bon flux de caixa. Això comporta que la seva situació financera i la seva existència es vegin amenaçades. Per exemple, una empresa que produeix abundantment, però que no ven prou i no és capaç de pagar als seus prestadors.
- Risc operatiu: Es l'exposició al risc per errors interns en qualsevol dels passos que es realitzen en l'operativa normal del negoci. Augmenta especialment quan hi ha manca de controls dins l'empresa, fallides tecnològiques, mala administració, errors humans, falta de personal...
- Risc de crèdit: És la probabilitat de que una de les dues parts d'un contracte creditici no assumeixi les seves obligacions de pagament per insolvència o qualsevol altre

motiu. El següent apartat i tota la resta del projecte està dedicat a explicar-lo més profundament.

També existeixen altres tipus de riscos financers com ara el *risc regulatori*, que seria el risc de que alguna variació en la regulació afecti negativament al negoci; el *risc de model*, que és el risc de que hi hagi defectes en algun model relacionat amb el *pricing*, avaluació del risc o negoci, i que porti a una presa de decisions incorrecta; o bé el *risc reputacional*, que és el perill de que una opinió pública negativa impedeixi o disminueixi la capacitat d'una entitat financera per a realitzar els seus negocis.

És important saber identificar en quina situació i risc es troba una empresa o individu en cada moment, i avaluar-ne constantment els possibles impactes en cas desfavorable. Si no, no es podrà reaccionar a temps ni prevenir qualsevol esdeveniment que podria malmetre la imatge de l'empresa, donar avantatge a la competència o fins i tot haver de declarar la fallida.

1.2 Risc de crèdit

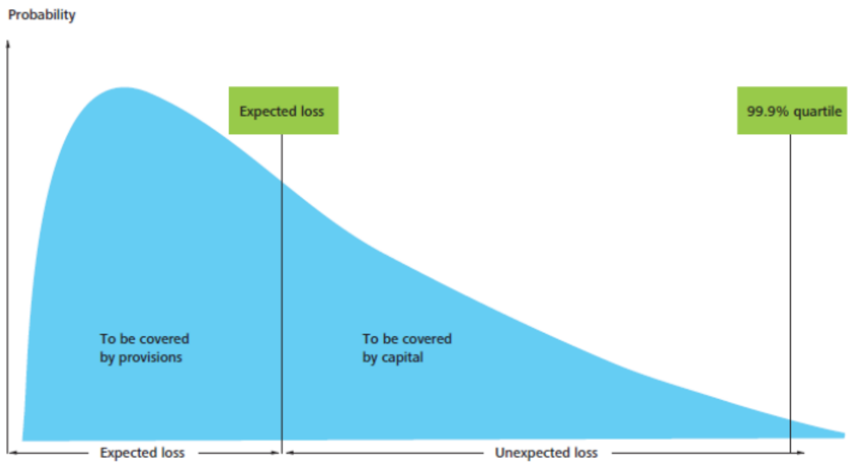
Aquest tipus de risc sorgeix de la possibilitat de que una de les parts del contracte de l'instrument financer, normalment els prestataris (ja siguin clients, empreses o fins i tot governs de països), incompleixin les seves obligacions de pagament per motius d'insolvència o per la simple incapacitat de pagament. Això produiria pèrdues financeres per la part dels prestadors, i problemes exponencialment complexos que en derivin.

Per a evitar-ho, els governs dels països d'arreu del món posen unes normes als bancs en relació a l'oferta de préstecs. L'entitat financera prestadora està obligada a cobrir les seves pèrdues (esperades i inesperades) derivades del risc de crèdit, subjectes a una certa normativa, ja que sinó les conseqüències de que caiguessin en fallida no només serien negatives per a els propis bancs sinó que ho serien per a una gran part de la població del país. Per exemple, en el curs d'un any, es calcula la pèrdua esperada, que és aquella que l'entitat espera perdre segons el curs normal de tots els seus negocis, i ha de reservar unes *provisions* per a cobrir-la. En canvi, la pèrdua inesperada és la quantitat que l'entitat podria perdre per raons fora del curs normal del negoci, i generalment és molt més difícil de predir que la pèrdua esperada. Per a cobrir aquesta pèrdua, l'entitat ha de reservar un *capital*.

La gestió del risc de crèdit es pot dividir en etapes de la següent manera:

1. *Admissió*: Existeix un sistema de qualificació creditícia que s'encarrega de donar un *Rating* i un *Scoring* (s'expliquen més endavant) per a cada client.
2. *Seguiment*:
   (a) Càlcul de la pèrdua esperada (provisions).
   (b) Estimació de la pèrdua inesperada (capital).

La imatge següent mostra la distribució de les pèrdues d'una entitat financera prestadora:



La pèrdua esperada (PE) es calcula amb la fórmula següent:

$$PE = PD * EAD * LGD$$

I la pèrdua inesperada (PI) es calcula amb la fórmula:

$$PI = LGD * EAD * f(PD)$$

On la probabilitat d'incompliment (PD) o *probability of default* recull la probabilitat de que el client no compleixi amb les seves obligacions contractuals. L'EAD, que és l'*exposure at default* fa referència a l'exposició a l'impagament que està definida com l'import del deute restant en el moment exacte que el prestatari deixa de pagar. La LGD és la *loss given default* i és el rati de pèrdua en cas d'incompliment, un percentatge d'un crèdit que, una vegada impagat i efectuades totes les gestions per a poder tornar-lo a temps, resulta finalment incobrable.

Aquestes variables (la PE i la PI) són molt importants per a totes aquelles entitats que ofereixen préstecs, béns o qualsevol altre producte a canvi d'una retribució major, tot i que sigui amb un percentatge d'interès molt baix, a la que s'havia donat inicialment. Per a aquestes és important saber gestionar de forma adequada les incerteses que es generen i, per tant, quantificar quant risc creditici té cada client per tal d'evitar al màxim possible tenir pèrdues. També és important per a les entitats tenir una mesura d'aquestes possibles pèrdues i tenir un fons de reserva de capital per si de cas no es retornessin aquests crèdits. Aquestes provisions no poden ser ni massa elevades, ja que aquests diners quiets podrien estar generant-ne més si estiguessin invertits, ni massa escasses, ja que llavors l'entitat es veuria molt compromesa econòmicament en cas d'impagament. Han de ser d'una quantitat justa, han de fer front als impagaments esperats durant un període, normalment d'un any, de forma que no causin danys en els seus comptes, ajustant-la tant com es pugui.

Ara bé, no existeix una única forma de risc, sinó que n'hi ha més, per exemple, el *risc col·lateral*, que depèn de les garanties existents del deute. També existeix el *risc de concentració*, que és present quan s'entreguen quantitats copioses o excessives a pocs prestataris que provoca un pic de pèrdues anuals, en el supòsit que aquests acabin essent morosos. A més, ja s'ha explicat a la introducció que a cada persona i empresa se li assigna un risc depenent del seu perfil. Per tant, la dificultat de modelitzar el risc sempre ha estat molt present ja que hi ha moltes variables en joc.

Va ser a partir de la greu crisi financera global de 2007-2008 que multituds de bancs i empreses van declarar-se en fallida degut a la gran quantitat de préstecs (en concret crèdits hipotecaris) que havien concedit a clients amb elevat risc creditici. Òbviament, això va comportar que un gran nombre d'empreses es declaressin en fallida i que hi hagués una urgent conscienciació en aquest sentit dels governs i, en concret, de les entitats financeres. A partir de llavors, han volgut entendre i gestionar molt millor el risc creditici, i la gran majoria de bancs tenen un departament de *Credit Risk Management*, o bé subcontracten consultories especialitzades, ja que aquestes no només tenen departaments concrets, sinó que al ser contractades per altres bancs, la competitivitat entre aquests està assegurada.

Sense una avaluació íntegra del risc de crèdit, els bancs no tenen altra forma de saber si les reserves de capital reflecteixen amb tota seguretat els riscos presents. Per exemple, si un banc presenta una reserva de capital i provisions insuficient llavors no serà capaç de fer front a les despeses causades pels impagaments, mentre que si és massa elevada aleshores perdrà competitivitat per no haver invertit tant com la resta d'entitats bancàries, a més de l'esforç que suposa augmentar aquesta reserva. Es busca llavors un punt mig entre les dues, amb cert marge conservador, tant ajustat com sigui possible, fent servir models per a cada variable del risc de crèdit.

Un *model* no deixa de ser un mètode, un sistema o un enfocament quantitatiu que aplica teories, tècniques o suposicions ja siguin estadístiques, econòmiques, financeres o matemàtiques, per a processar dades i extreure'n estimacions quantitatives. Les entitats financeres utilitzen aquests models per a la estimació i posterior gestió del risc de crèdit.

Per a que no succeeixin esdeveniments negatius com ara el no retorn d'un préstec, i per a garantir que les reserves de capital reflecteixin el perfil de risc apropiadament, s'ha d'implementar una solució integrada i quantitativa que no afecti en el rendiment del banc, és a dir, un model. Aquest ha d'incloure:

1. Una gestió perfecta de l'etapa de modelització.
2. Possibilitat d'avaluació i monitorització en temps real.
3. Comprovació amb dades anteriors de que el model funcioni segons els criteris establerts.
4. Fer servir eines de visualització de dades i d'intel·ligència de negoci que permetin extreure la informació imprescindible de cada cas.

S'ha de tenir en compte que el procés serà diferent si s'està analitzant el risc per a grans empreses i empreses corporatives o si bé són particulars o petites empreses. En el cas de grans empreses, un conjunt d'experts en mesura del risc d'impagament fa un anàlisi individualitzat i en conclou el risc i l'interès. En el cas de petites empreses o particulars, es construeixen models que avaluen les operacions de forma massiva a partir dels impagaments històrics que ha tingut l'entitat financera. Dins d'aquest concepte n'hi ha d'altres com la puntuació i la classificació de crèdit, que s'expliquen la secció següent.

1.3 Probabilitat d'impagament (PD)

Com s'ha descrit anteriorment, la probabilitat d'impagament és una mesura de qualificació creditícia que se li dóna a cada client segons les seves característiques i de la situació econòmica global que es viu en aquell precís moment. L'objectiu és estimar quant de probable és que acabin no pagant, durant un cert període de temps, que normalment és un any.

Quan arriba un client que desitja un préstec de l'entitat financera, el banc estudia el seu cas i determina la seva PD (recordem, *Probability of Default*, probabilitat d'impagament). En el supòsit que presenti un valor suficientment elevat, s'aplicaran interessos importants en el préstec i proporcionalment es reservarà més capital per assegurar-ne la seva viabilitat.

Per a orientar el valor de la PD, s'utilitzen dues eines anomenades puntuació de crèdit, o en anglès *Scoring*, que qualifica alguna de les contraparts d'un contracte (normalment usat per particulars i petits negocis, on el titular pot comportar-se d'una forma diferent segons el tipus de producte), i amb aquesta es fa una classificació de crèdit, o *Rating*, que normalment avalua el binomi client-operació (usualment es fa servir per a classificar grans empreses amb un gran nombre d'operacions contractuals). Es basen en l'ús d'instruments estadístics focalitzats en l'estimació de la probabilitat d'impagament per a després decidir si el possible prestatari acaba rebent el crèdit o no. La puntuació de crèdit impacta en una multitud de transaccions financeres que inclouen les hipoteques, les targetes de crèdit o els préstecs privats.

Popularment es fan servir dues maneres per a obtenir aquestes puntuacions. La més utilitzada és la *Puntuació FICO®* (Fair Isaac COrporation), que el fa servir el 90% del sector, i l'altra és la *Puntuació Vantage*. Aquestes dues metodologies són molt semblants, si bé les dues poden prendre valors entre 300 i 850, la única diferència destacable és la classificació de l'individu dins d'aquests valors, essent la FICO® més restrictiva. Tot i això, la puntuació mínima i màxima que es pot obtenir és la mateixa en ambdós casos.

Aquestes puntuacions no només serveixen per a decidir si concedir un préstec o amb quines condicions (interès, nombre de terminis, quantitat...), sinó que també s'utilitza per a l'obtenció de targetes de crèdit, per exemple. Els individus i empreses disposen d'un informe de puntuació gratuït a l'any, i és molt recomanable tenir-lo actualitzat anualment.

La puntuació de crèdit individual s'obté tenint en compte els següents factors:

- Variables d'operació
  - Procedència de l'operació
  - Destí
  - Import del crèdit
  - Quota
  - Terminis
- Variables de perfil socio-econòmic:
  - Edat
  - Estat civil
  - Ingressos i despeses
  - Situació laboral i sector
  - Antiguitat
- Variables de vinculació a l'entitat prestadora
  - Antiguitat com a client

  - Domiciliació de la nòmina i rebuts
  - Saldo

- Experiències de pagament anteriors:
  - S'acut a ASNEF (Associació Nacional d'Establiments Financers de Crèdit - Fitxer de Morosos) i EXPERIAN (Fitxers de solvència)
  - Comportament intern amb l'entitat.

També cal comentar que els que proporcionen els valors d'aquests termes no són ni *FICO*, ni *Vantage*, ni tampoc els departaments de risc de crèdit dels bancs, sinó un altre tipus de companyies que recopilen aquestes dades relacionades amb els crèdits. Les tres principals són *Experian, Equifax i TransUnion*, i degut a que les dades de cada companyia poden ser diferents, les puntuacions també poden variar segons la que s'esculli.

Podem guiar-nos en la classificació (*Rating*) segons la puntuació (*Scoring*) de cada metodologia (*FICO* o *Vantage*) amb la taula següent:

Rating	Escala de FICO	Escala de Vantage
Excel·lent	800-850	781-850
Molt Bona / Bona	740-799	661-780
Bona / Justa	670-739	601-660
Justa / Pobre	580-669	501-600
Pobre / Molt Pobre	300-579	300-500

Taula 1. *Rating* segons l'*Scoring*

Com es pot veure, realment la funció de l'*Scoring* és proporcionar una classificació objectiva segons el *Rating*, amb uns termes molt entenedors, però no proporciona cap estimació de la probabilitat d'impagament. A més, el *Rating* fa una valoració del binomi client-transacció (proporciona un valor per a cada transacció de cada client), i l'*Scoring* avalua clients en particular. Existeixen noves tecnologies i programaris que proporcionen altres metodologies més avançades i precises per a modelitzar el risc de crèdit, però només les tenen uns pocs experts als països que es poden permetre invertir més en el sector bancari. De totes maneres, hi ha un capítol dedicat als models amb algoritmes de *Machine Learning*, que és una tecnologia relativament nova i el seu us no és tant inassumible per a empreses tradicionals.

L'avaluació dels crèdits i la valoració per a grans empreses i corporacions, inclús de governs d'un país, la fan normalment agències dedicades a la qualificació de crèdit. El que fan és mesurar la solvència de les entitats, és a dir les capacitats per afrontar les seves obligacions financeres. Si la qualificació és prou elevada voldrà dir que hi ha una alta probabilitat que l'empresa torni el préstec en la seva totalitat sense problemes, mentre que si és baixa és probablement perquè ha tingut problemes per a pagar en el passat i que podria tornar-ho a fer, o per algun altre motiu dels que s'han enumerat abans.

Existeixen unes ràtios amb les quals aquestes agències de qualificació de crèdit calculen, mitjançant fórmules matemàtiques, el *Rating* per a empreses. Aquest ràtios es poden agrupar de la següent manera:

- Ràtios de rendibilitat: mesuren el nivell de rendibilitat dels recursos financers dels que disposa l'empresa i de les inversions realitzades.

- Ràtios d'estructura financera: mesuren el grau de deute de l'empresa, la seva autonomia financera i l'equilibri entre les masses patrimonials més importants.

- Ràtios de liquiditat: mesuren el grau de liquiditat dels actius i la seva capacitat per a fer front a les obligacions a curt termini, així com la capacitat de generar flux de caixa per a l'empresa.

- Ràtios de rotació: mesuren l'eficiència amb la que l'empresa realitza les seves activitats.

- Ràtios de costes: són un conjunt d'indicadors que permeten conèixer el pes dels costos ordinaris i financers en el conjunt de l'activitat de l'empresa.

Així doncs, el *Rating* que obté una empresa o individu té un pes molt important per a l'obtenció del préstec. En la taula següent veiem que la determinació del *Credit Rating* es fa mitjançant una escala alfabètica de valoracions, elaborada per les agències de qualificació de risc de crèdit, en aquest cas, *Moody's*:

Aaa	Lowest level of credit risk
Aa1	Very low credit risk
Aa2	
Aa3	
A1	Low credit risk
A2	
A3	
Baa1	Moderate credit risk
Baa2	
Baa3	
Ba1	Substantial credit risk
Ba2	
Ba3	
B1	High credit risk
B2	
B3	
Caa1	Very high credit risk
Caa2	
Caa3	
C	Extremely speculative

Taula 2. Sistema alfanumèric que utilitza l'agència *Moody's*

També n'hi ha dues més, *S&P*(*Standard and Poor's*) i *Fitch*, que s'utilitzen molt (entre les tres tenen el 95% d'aquest mercat). Cada agència fa servir la seva escala d'etiquetes de classificació, tot i que totes són alfanumèriques i semblants.

Tal i com es pot observar, *Moody's* utilitza l'etiqueta Aaa per a distingir aquelles empreses que presenten el mínim risc possible, i en canvi *S&P* i *Fitch* fan servir l'etiqueta AAA com a excel·lència. En canvi, el C seria el grup que presenta el major risc creditici. Una de les consideracions a tenir en compte és que si el *Rating* es igual o inferior a Ba1 en l'escala de *Moody's* es considera un *Junk bond*, mentre que per sobre d'aquesta classificació es considera un *Investment bond.* També es podria considerar com a pitjor rating tenir una D, que seria estar ja en *default*, però no es posa perquè el risc d'impagament òbviament és 1.

Per a situar alguns exemples pràctics, tots segons l'escala de *Moody's*, podem fixar-nos, per exemple, en que el govern espanyol té una qualificació de Baa2, l'Alemany té Aaa i el Francès Aa1.

També cal comentar que aquestes empreses de qualificació sempre han generat una certa desconfiança a l'hora de qualificar grans empreses i governs estatals, ja que són aquests mateixos els que paguen a les agències per a que analitzin la seva solvència, i per tant hi ha gent que qüestiona la independència total per a realitzar l'anàlisi.

Una altra curiositat és que aquestes tres agències de qualificació de risc creditici són propietat d'empreses privades, i en cap cas són finançades amb fons públics.

En resum, el *Rating* i l'*Scoring* serveixen per a classificar a un client i, un cop ja classificat, se li assigna una probabilitat d'entrar en impagament, que és la PD, segons la puntuació rebuda, que en el cas de particulars o de PYMEs és numèrica i en el cas de grans empreses i empreses corporatives, és alfanumèrica.

Per tancar aquesta secció, cal comentar que una qualificació/*Rating*/*Scoring* no significa res per ella mateixa, sinó que és necessària la comparació amb una altra d'una altra contrapart. Mitjançant processos anomenats *calibrats*, cada qualificació porta associada una *PD*, que és un concepte estadístic que sí que te sentit per ell mateix.

1.4 Exposició en el moment d'incompliment (EAD)

L'exposició en el moment d'incompliment, o *Exposure at Default* (EAD) és un factor important per al càlcul de la pèrdua esperada, que ve definida com l'import del deute pendent de pagar en el moment que es produeix l'esdeveniment. Aquesta pèrdua depèn directament de la quantitat amb la qual el banc o entitat financera està exposada respecte al prestatari en el moment de l'impagament, ja que l'incompliment tindrà lloc en una data futura desconeguda.

Els bancs calculen l'EAD per cadascun dels crèdits disponibles per a determinar el risc conjunt d'impagament. Es tracta d'un nombre variable segons si el prestatari torna els diners corresponents o no. El que s'acaba mesurant és l'extensió en la qual el banc es veu exposat respecte el client en el cas que, passat un temps, aquest no pagui.

L'EAD pot calcular-se de la següent manera:

$$EAD = Disposat + Disponible * CCF$$

On *Disposat* és la quantitat que s'ha fet servir del límit del crèdit, *Disponible* és aquell import restant que el prestatari encarà té pendent consumir, i *CCF* és el *Credit Conversor Factor*, és a dir el rati percentual entre la quantitat monetària que el client demana d'un nou crèdit i el capital restant del préstec actual. Un exemple seria que si una persona posseeix un crèdit al que li resten 20.000€ disponibles i en demana un altre de 5.000€, llavors el *CCF* resultant és un 25%.

1.5 Severitat (LGD)

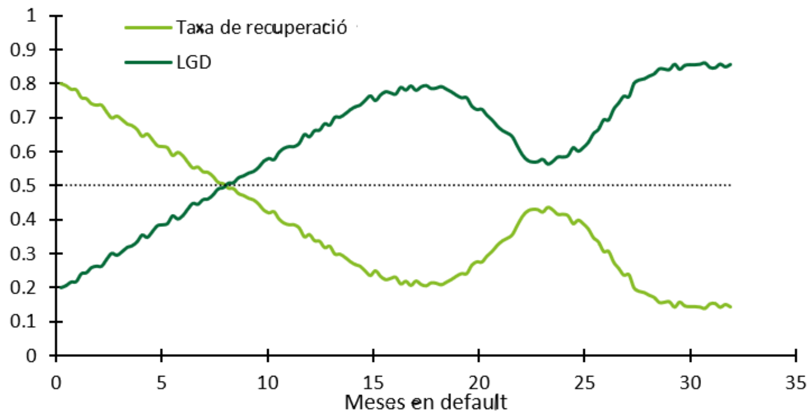
La severitat o *Loss Given Default* (LGD) és una mètrica imprescindible en el risc de crèdit. És la quantitat de diners que el banc o l'entitat financera prestadora perd en el moment que el prestatari incompleix amb el pagament, i normalment està expressada en forma percentual sobre el total de l'exposició.

Conceptualment, és molt fàcil d'entendre com el complementari de la *Taxa de recuperació (TR)*:

$$LGD = 1 - Taxa\ de\ recuperació\ (TR)$$

La *Taxa de recuperació* és la mesura de quant import es pot recuperar del deute en impagament.

I gràficament:



Gràfic de la proporcionalitat entre LGD i TR.

La pèrdua LGD es sol determinar analitzant altres impagaments. La quantificació de les pèrdues és complexa i és necessari analitzar un ampli conjunt de variables que, després d'un procés exhaustiu, permeten determinar la severitat, com ara el tipus d'instrument financer, possibles problemes corporatius o condicions macroeconòmiques.

Per exemple, si un banc ofereix un milió d'euros a una empresa i aquesta, per causes qualssevol, acaba no pagant, la pèrdua per al banc no és necessàriament d'un milió d'euros. Existeixen altres factors que l'acabaran determinant, com ara la quantitat de béns dels que disposa l'empresa en forma de garantia del contracte, o si ja s'havien pagat unes quantes quotes prèviament. Aquest factors reduirien substancialment la pèrdua inicial del capital atorgat a l'empresa.

Per altra banda, també augmenten el valor de la LGD tots els tràmits i gestions que realitzen les entitats prestadores davant els impagaments, com ara l'enviament de cartes, els costos judicials, advocats, o empleats destinats a la gestió de la morositat.

També es pot produir una altra situació, que és que el client no pagui però que en un termini raonable retorni els rebuts ja vençuts i es posi al dia del contracte. En aquest cas, la severitat seria un percentatge molt baix, proper a 0, donat que es considerarien com a pèrdues només les gestions dels bancs davant la morositat, que acabem de mencionar.

La LGD i l'EAD són dos factors molt semblants. La principal diferència és que la LGD inclou la possibilitat de recuperar capital monetari de l'incompliment. En el cas que el prestatari incompleixi el pagament d'una hipoteca després d'uns quants anys pagant mensualment, l'EAD seria la quantitat que li quedava per a liquidar-la completament. Si el banc vengués l'habitatge, recuperaria una certa quantitat de l'EAD, que sí que estaria inclosa en el càlcul de la LGD.

Veurem ara un exemple que té en compte tots els factors vists fins ara.

Exemple 1.1. *Imaginem una parella que demana una quantitat de 225.000€ per un préstec d'un habitatge. Després d'estar pagant quotes durant varis anys, la parella queda estancada financerament parlant, i tenen problemes per a seguir amb els pagaments mensuals fins que acaben incomplint amb el pagament acordat.*

Posem que en aquest moment, l'EAD és de 186.500€. Vist això, el banc es fa amb el control del condomini i pot vendre'l per 160.000€, obtenint així unes pèrdues netes de 26.500€ i un LGD del 14.21%, ja que:

$$P\grave{e}rdues\ netes = 186.500€ - 160.000€ = 26.500€$$

$$LGD = \frac{186.500€ - 160.000€}{186.500€} = 0.14209 \approx 14.21\%$$

En aquesta mateixa situació, la pèrdua esperada es calcula tenint en compte els següents factors: un LGD del 14.21%, una probabilitat d'incompliment (PD) del 100% (ja que se sap que van incomplir) i una EAD de 186.500€, obtenim llavors:

$$P\grave{e}rdua\ esperada\ (PE) = 1 * 0.14209 * 186.500€ = 26.500€$$

Aquesta pèrdua esperada s'ha calculat a partir d'un cas on es coneix tot ja que ja ha passat, però podria ser que l'entitat financera estigui interessada en conèixer les pèrdues potencials que podrien patir, sense que fos segur que acabi ocorrent. Per a fer això, el banc assumeix que la probabilitat d'impagament (PD) de la parella no és del 100% sinó del 60%. D'aquesta manera la nova pèrdua esperada seria de 15.900€, ja que:

$$P\grave{e}rdua\ esperada\ (PE) = 0.6 * 0.14209 * 186.500€ = 15.900€$$

Capítol 2

Introducció al càlcul estocàstic

En aquest capítol, s'introdueixen les eines matemàtiques necessàries per a modelitzar actius financers i opcions de compra i venda. En temps continu, els aspectes tècnics són més complexos i també més difícils de desenvolupar que a temps discret, però coincideixen en la majoria d'idees bàsiques.

Perquè considerem models a temps continu? La motivació principal la tenim en la naturalesa dels processos que volem modelitzar. En la pràctica, el flux dels preus de mercat varia tant que inclús amb models de temps continu és difícil de seguir. A més, els models de temps continu ens porten a càlculs més explícits, i fins i tot hem de recórrer a mètodes numèrics. De fet, el model a temps continu més usat i conegut globalment és el *model de Black-Scholes (Fisher Black i Myron Scholes)*, que acaba portant a una equació relativament senzilla.

2.1 Comentaris generals sobre processos a temps continu

Es comença amb unes definicions matemàtiques per a entendre millor els models a temps continu, com ara la filtració natural, el *Moviment Brownià* o les martingales a temps continu.

Però abans de començar, pensem, què vol dir exactament que un procés sigui a temps continu?

Definició 2.1. *Sigui $(E, \mathscr{E})$ un espai mesurable. Un procés estocàstic a temps continu amb espai d'estats $(E, \mathscr{E})$ és una família $(X_t)_{t \in \mathbb{R}^+}$ de variables aleatòries definides en un espai de probabilitat $(\Omega, \mathscr{F}, \mathbb{P})$ que pren valors en $(E, \mathscr{E})$.*

Remarca 2.1.

- *En la pràctica,* t *és el temps.*
- *Un procés pot també ser considerat una variable aleatòria: per a cada $\omega \in \Omega$ tenim associada una aplicació de $\mathbb{R}^+$ a $E: \ t \to X_t(\omega)$, anomenats* camins *o trajectòries del procés.*
- *Només es treballarà amb processos en què el temps estigui indexat en un interval finit $[0, T]$.*

Definició 2.2. *Una filtració d'un espai de probabilitat $(\Omega, \mathscr{F}, \mathbb{P})$ és una família creixent $(\mathscr{F}_t)_{t \geq 0}$ de σ-àlgebres incloses en $\mathscr{F}$.*

La σ-àlgebra $\mathscr{F}_t$ representa la informació disponible a l'instant t. Es diu que un procés $(X_t)_{t\geq 0}$ és adaptat si, per tot t, X_t és $\mathscr{F}_t$-mesurable.

Definició 2.3. *Considerem un espai de probabilitat $(\Omega, \mathscr{F}, \mathbb{P})$. Direm que un conjunt $A' \subset \Omega$ (que pot no pertànyer a $\mathscr{F}$) és negligible si existeix un $A \in \mathscr{F}$ tal que $A' \subset A$ i $\mathbb{P}(A) = 0$.*

Remarca 2.2. *D'ara endavant, es treballarà amb filtracions que tenen la propietat següent:*

Si A' $\subset A \in \mathscr{F}$ i si $\mathbb{P}(A) = 0$, llavors, per tot $t, A' \in \mathscr{F}_t$.

En altres paraules, $\mathscr{F}_t$ conté tots els conjunts de $\mathscr{F}$ amb probabilitat $\mathbb{P}$ nul·la (negligibles), i gràcies a això, es té que si X=Y quasi segurament respecte la mesura de probabilitat $\mathbb{P}$, i Y és $\mathscr{F}_t$-mesurable, llavors X és també $\mathscr{F}_t$-mesurable.

Es pot construir ara una filtració generada pel procés $(X_t)_{t\geq 0}$ i escrivim

$$\mathscr{F}_t = \sigma(X_s, s \leq t).$$

En general, aquesta filtració no satisfà la condició anterior. Tanmateix, si es canvia $\mathscr{F}_t$ per $\tilde{\mathscr{F}}_t$, que és la σ-àlgebra generada tant per $\mathscr{F}_t$ com per $\mathscr{N}$ (la σ-àlgebra generada per tots els conjunts negligibles de $\mathscr{F}$), s'obté la corresponent filtració que satisfà la condició de l'anterior remarca. S'anomena *filtració natural* del procés $(X_t)_{t\geq 0}$.

El concepte de *temps d'aturada* serà molt útil. Un *temps d'aturada* és una variable aleatòria que depèn del procés subjacent. En altres paraules, en un temps donat t, es sap si el *temps d'aturada* és menor o igual que t. La definició següent ho formalitza:

Definició 2.4. *El temps d'aturada respecte a la filtració $(\mathscr{F}_t)_{t\geq 0}$ és una variable aleatòria τ, amb valors en $\mathbb{R}^+ \cup \{+\infty\}$, tal que, per cada $t \geq 0$,*

$$\{\tau \leq t\} \in \mathscr{F}_t.$$

La σ-àlgebra associada amb τ està definida com

$$\mathscr{F}_t = \{A \in \mathscr{F}, \quad \forall t \geq 0, A \cap \{\tau \leq t\} \in \mathscr{F}_t\}$$

Aquesta σ-àlgebra representa la informació disponible abans del temps aleatori τ.

Proposició 2.1.

- *Si S és un temps d'aturada, llavors S és $\mathscr{F}_S$-mesurable.*
- *Si S és un temps d'aturada, finit quasi segurament, i $(X_t)_{t\geq 0}$ és un procés adaptat continu, aleshores X_S és $\mathscr{F}_S$-mesurable.*
- *Si S i T són dos temps d'aturada tals que $S \leq T$ q.s., aleshores $\mathscr{F}_S \subset \mathscr{F}_T$.*
- *Si S i T són dos temps d'aturada, aleshores $S \wedge T = \inf(S, T)$ és un temps d'aturada. En particular, si S és un temps d'aturada i t és un temps determinista, aleshores $S \wedge t$ és un temps d'aturada.*

Demostració. Veure *Lamberton, D; Lapeyre, B. (2008): Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall / CRC.* □

2.2 Moviment Brownià

Un exemple d'un procés estocàstic de particular importància és el *Moviment Brownià.* És bàsic per a entendre la majoria de models financers.

Definició 2.5. *Un Moviment Brownià és un procés estocàstic continu $(X_t)_{t\geq 0}$ a valors reals, amb increments estacionaris i independents. És a dir, compleix:*

- *continuïtat: l'aplicació $s \to X_s(\omega)$ és contínua $\mathbb{P}-q.s.$*
- *increments estacionaris: si $s \leq t$, $X_t - X_s$ i $X_{t-s} - X_0$ tenen la mateixa llei.*
- *increments independents: si $s \leq t$, $X_t - X_s$ és independent de*

$$\mathscr{F}_s = \sigma(X_u, u \leq s).$$

Teorema 2.1. *Si $(X_t)_{t\geq 0}$ és un Moviment Brownià, llavors $X_t - X_0$ és una variable aleatòria amb distribució normal, amb mitjana rt i variància $\sigma^2 t$, on r i σ són constants reals.*

Demostració. Veure *Gikham, I; Skorhod, A. V. (1969): Introduction to the theory of random processes. Traduït del Rus a l'anglès per Scripta Technica, Inc. W. B. Saunders Co.* □

Remarca 2.3. *Un Moviment Brownià és Estàndard si per tot t:*

$$X_0 = 0\, q.s. \quad \mathbb{E}(X_t) = 0, \quad \mathbb{E}(X_t^2) = t.$$

D'ara endavant, se suposa que, si no es diu el contrari, tots els Moviments Brownians seran Estàndard. La densitat de X_t és:

$$\frac{1}{\sqrt{2\pi t}} exp\left(-\frac{x^2}{2t}\right)$$

S'acaba d'observar que per qualsevol t, X_t és una variable aleatòria normal. Ara es veurà una definició útil per a arribar a un resultat més fort que el que s'acaba de veure:

Definició 2.6. *Una variable aleatòria amb valors en $\mathbb{R}^d$, $X = (X_1, \cdots, X_d)$, és un vector Gaussià si, per qualsevol col·lecció de nombres reals $a_1, \cdots, a_d$, la variable aleatòria escalar $\sum_{i=1}^d a_i X_i$ és Gaussiana.*

Teorema 2.2. *Sigui $(X_t)_{t\geq 0}$ un Moviment Brownià (Estàndard) i si $0 \leq t_1 < \cdots < t_n$, aleshores $(X_t, \cdots, X_{t_n})$ és un vector Gaussià.*

Demostració.
Considerem $0 \leq t_1 < \cdots < t_n$. Aleshores el vector aleatori $X = (X_{t_1}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}})$ està composat per variables aleatòries normals independents (pel *Teorema 2.1* i la definició de Moviment Brownià). Llavors, el vector X és Gaussià i també ho és $(X_t, \cdots, X_{t_n})$. □

També és necessària una definició de Moviment Brownià respecte de la filtració $(\mathscr{F}_t)$:

Definició 2.7. *Un procés estocàstic continu a valors reals és un $(\mathscr{F}_t)$-Moviment Brownià si satisfà:*

- *Per tot* $t \geq 0$, X_t *és* $(\mathscr{F}_t)$*-mesurable.*
- *Si* $s \leq t, X_t - X_s$ *és independent de la* σ*-àlgebra* $\mathscr{F}_s$.
- *Si* $s \leq t, X_t - X_s$ *i* $X_{t-s} - X_0$ *tenen la mateixa llei.*

Remarca 2.4. *El primer d'aquests punts mostra que* $\sigma(X_u, u \leq t) \subset \mathscr{F}_t$. *A més, és fàcil de comprovar que un* $(\mathscr{F}_t)$*-Moviment Brownià és també un Moviment Brownià respecte la seva filtració natural.*

2.3 Martingales a temps continu

El concepte de martingala és crucial per a explicar la idea d'arbitratge. La definició següent és una extensió de la de temps discret, ja feta a l'assignatura de Modelització:

Definició 2.8. *Es considera un espai de probabilitat* $(\Omega, \mathscr{F}, \mathbb{P})$, *i una filtració* $(\mathscr{F}_t)_{t\geq 0}$ *en aquest espai. Una família adaptada de variables aleatòries* $(M_t)_{t\geq 0}$ *tal que* $\mathbb{E}(|M_t|) < \infty$, *per tot* t, *és:*

- *una martingala si, per tot* $s \leq t$, $E(M_t|\mathscr{F}_s) = M_s$;
- *una supermartingala si, per tot* $s \leq t$, $E(M_t|\mathscr{F}_s) \leq M_s$;
- *una submartingala si, per tot* $s \leq t$, $E(M_t|\mathscr{F}_s) \geq M_s$.

Remarca 2.5. *D'aquesta definició es segueix que si* $(M_t)_{t\geq 0}$ *és una martingala, llavors* $\mathbb{E}(M_t) = \mathbb{E}(M_0)$, *per tot* t.

Una opció de compra o venda pot ser exercida en qualsevol moment abans de la maduresa T. La decisió d'exercir o no al temps n es fa amb la informació disponible al temps n. En un model de temps continu construït en un espai de probabilitat, filtrat i finit, $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{0\leq n\leq N}, \mathbb{P})$, la data en la que s'exerceix l'opció ve donada per una variable aleatòria anomenada *temps d'aturada.*

Definició 2.9. *Una variable aleatòria* ν *que pren valors en* $\{0, 1, 2, \cdots, N\}$ *és un temps d'aturada si, per qualsevol* $n \in \{0, 1, 2, \cdots ., N\}$,

$$\{\nu = n\} \in \mathscr{F}_n.$$

Farem servir una definició equivalent per a generalitzar el concepte de *temps d'aturada* a una configuració de temps continu.

Definició 2.10. *Sigui* $(X_n)_{0\leq n\leq N}$ *una successió adaptada a la filtració* $(\mathscr{F}_n)_{0\leq n\leq N}$ *i sigui* ν *un temps d'aturada. La successió aturada a temps* ν *està definida com:*

$$X_n^\nu(\omega) = X_{\nu(\omega)\wedge n(\omega)},$$

és a dir, en el conjunt $\nu = j$, *tenim:*

$$X_n^\nu = \begin{cases} X_j & si\, j \leq n \\ X_n & si\, j > n \end{cases}$$

Notem que $X_N^\nu(\omega) = X_{\nu(\omega)}(\omega)(= X_j$ *en* $\{\nu = j\}$.

Abans de continuar, veiem una definició i proposició per a veure què és una successió previsible i què és una *transformació de martingala*:

Definició 2.11. *Una successió adaptada $(H_n)_{0\leq n\leq N}$ de variables aleatòries és previsible si, per tot $n \geq 1$, H_n és $\mathscr{F}_{n-1}$-mesurable.*

Proposició 2.2. *Sigui $(M_n)_{0\leq n\leq N}$ una martingala i $(H_n)_{0\leq n\leq N}$ una successió previsible respecte la filtració $(\mathscr{F}_n)_{0\leq n\leq N}$. Denotem $\Delta M_n = M_n - M_{n-1}$. La successió $(X_n)_{0\leq n\leq N}$ definida per*

$$X_0 = H_0 M_0$$
$$X_n = H_0 M_0 + H_1 \Delta M_1 + \cdots + H_n \Delta M_n, \quad \text{per } n \geq 1$$

és una martingala respecte de la filtració $(\mathscr{F}_n)_{0\leq n\leq N}$.

Aquesta (X_n) s'anomena transformació d'una martingala per (H_n)

Demostració. Veure *Lamberton, D; Lapeyre, B. (2008): Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall / CRC.* □

Proposició 2.3. *Sigui (X_n) una successió adaptada i ν un temps d'aturada. La successió aturada $(X_n^\nu)_{0\leq n\leq N}$ és adaptada. A més, si (X_n) és una martingala (o una supermartingala o submartingala), aleshores (X_n^ν) és una martingala (o una supermartingala o submartingala).*

Demostració. Veiem que, per a $n \geq 1$, tenim:

$$X_{\nu\wedge n} = X_0 + \sum_{j=i}^{n} \phi_j (X_j - X_{j-1}),$$

on $\phi_j = \mathbf{1}_{\{j\leq\nu\}}$. Com que $\{j \leq \nu\}$ és el conjunt complementari de $\{\nu < j\} = \{\nu \leq j-1\}$, el procés $(\phi_n)_{0\leq n\leq N}$ és previsible. Aleshores és clar que $(X_{\nu\wedge n})_{0\leq n\leq N}$ és adaptat a la filtració $(\mathscr{F}_n)_{0\leq n\leq N}$.

A més, si (X_n) és una martingala, $(X_{\nu\wedge n})$ és també una martingala respecte $(\mathscr{F}_n)$, ja que és la transformació de martingala de (X_n). Anàlogament, es pot veure que si la successió (X_n) és una supermartingala (o una submartingala), la successió aturada és també una supermartingala (o una submartingala), fent servir la previsibilitat i la no negativitat de $(\phi_j)_{0\leq j\leq N}$. □

Proposició 2.4. *Si $(X_t)_{t\geq 0}$ és un $\mathscr{F}_t$-moviment Brownià Estàndard, llavors:*

1. *X_t és una $\mathscr{F}_t$-martingala.*
2. *$X_t^2 - t$ és una $\mathscr{F}_t$-martingala.*
3. *$exp(\sigma X_t - (\sigma^2/2)t)$ és una $\mathscr{F}_t$-martingala, per tot $\sigma \in \mathbb{R}$.*

Demostració.
Si $s \leq t$, llavors $X_t - X_s$ és independent de la σ-àlgebra $\mathscr{F}_s$. Llavors $\mathbb{E}(X_t - X_s|\mathscr{F}_s) = \mathbb{E}(X_t - X_s)$. Com que un moviment Brownià Estàndard té l'esperança igual a zero, tenim $\mathbb{E}(X_t - X_s) = 0$. Aleshores, *1.* ja està provat.

Per al segon punt, observem:

$$\mathbb{E}(X_t^2 - X_s^2|\mathscr{F}_s) = \mathbb{E}((X_t - X_s)^2 + 2X_s(X_t - X_s)|\mathscr{F}_s) = \mathbb{E}((X_t - X_s)^2|\mathscr{F}_s) + 2X_s\mathbb{E}(X_t - X_s)|\mathscr{F}_s),$$

i com que $(X_t)_{t\geq 0}$ té increments independents, $\mathbb{E}((X_t - X_s)^2|\mathscr{F}_s) = 0$, llavors:

$$\mathbb{E}(X_t^2 - X_s^2|\mathscr{F}_s) = \mathbb{E}((X_t - X_s)^2|\mathscr{F}_s)$$

A més, com que un moviment Brownià té increments independents i estacionaris, es segueix que

$$\mathbb{E}(X_t^2 - X_s^2|\mathscr{F}_s) = \mathbb{E}(X_{t-s}^2) = t - s.$$

L'última igualtat ve donada pel fet que X_t té una distribució normal amb mitjana zero i variància t. Això porta a que $\mathbb{E}(X_t^2 - t|\mathscr{F}_s) = X_s^2 - s$, si $s < t$.

Per últim, recordant que si g és una variable aleatòria normal estàndard, es té:

$$\mathbb{E}(e^{\lambda g}) = \int_{-\infty}^{+\infty} e^{\lambda x} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = e^{\lambda^2/2}.$$

Per altra banda, si $s < t$,

$$E(e^{\sigma X_t - \sigma^2 t/2}|\mathscr{F}_s) = e^{\sigma X_s - \sigma^2 t/2} E(e^{\sigma(X_t - X_s)}|\mathscr{F}_s)$$

perquè X_s és $\mathscr{F}_s$-mesurable. Com que $X_t - X_s$ és independent de $\mathscr{F}_s$, resulta:

$$E(e^{\sigma(X_t-X_s)}|\mathscr{F}_s) = \mathbb{E}(e^{\sigma(X_t-X_s)}) = \mathbb{E}(e^{\sigma X_{t-s}}) = \mathbb{E}(e^{\sigma g\sqrt{t-s}}) = exp\left(\frac{1}{2}\sigma^2(t-s)\right).$$

Això completa la prova. □

Si $(M_t)_{t\geq 0}$ és una martingala, la propietat $E(M_t|\mathscr{F}_s) = M_s$ és també certa si t i s són temps d'aturada *acotats*. Aquest resultat és una adaptació del *teorema de Mostreig Opcional*, que diu que sota certes condicions, l'esperança d'una martingala en un temps d'aturada és igual al seu valor esperat inicialment. Formalment:

Teorema 2.3. ***(Mostreig opcional)*** *Si $(M_t)_{t\geq 0}$ és una martingala a temps continu respecte de la filtració $(\mathscr{F}_t)_{t\geq 0}$, i si τ_1 i τ_2 són dos temps d'aturada tals que $\tau_1 \leq \tau_2 \leq K$, on* K *és un nombre real finit, llavors M_{τ_2} és integrable i:*

$$E(M_{\tau_2}|\mathscr{F}_{\tau_1}) = M_{\tau_1} \quad \mathbb{P} - q.s.$$

Demostració. Veure *Karatzas, I; Shreve, S. (1988): Brownian Motion and Stochastic Calculus. Springer.* □

Remarca 2.6.

- *Aquest resultat implica que si τ és un temps d'aturada acotat, aleshores $E(M_\tau) = E(M_0)$ (aplicant el teorema per a $\tau_1 = 0$ i $\tau_1 = \tau$ i prenent esperances.*

- *Si M_t és una submartingala, aleshores el mateix teorema és cert només variant l'igualtat anterior per:*

$$E(M_{\tau_2}|\mathscr{F}_{\tau_1}) \geq M_{\tau_1} \quad \mathbb{P} - q.s.$$

A continuació s'aplica aquest resultat per a estudiar les propietats del *temps d'arribada* a un cert nivell en un Moviment Brownià:

Proposició 2.5. *Sigui $(X_t)_{t\geq 0}$ un $(\mathscr{F}_t)$-Moviment Brownià. Per a cada nombre real* a, *tenim $T_a = \inf\{s \geq 0, X_s = a\}$ o $+\infty$ si aquest conjunt és buit. Aleshores, T_a és un temps d'aturada, finit quasi segurament, i la seva distribució està caracteritzada per la seva transformada de Laplace,*

$$\mathbb{E}(e^{-\lambda T_a}) = e^{-\sqrt{2\lambda}|a|}, \quad \lambda \geq 0.$$

Demostració.

- Per casos, primer suposem $a \geq 0$. Primer veurem que T_a és un temps d'aturada. De fet, com que X_s és continu,

  $$\{T_a \leq t\} = \bigcap_{\epsilon \in \mathbb{Q}^{+*}} \left\{ \sup_{s \leq t} X_s > a - \epsilon \right\} = \bigcap_{\epsilon \in \mathbb{Q}^{+*}} \bigcap_{s \in \mathbb{Q}^+,\, s \leq t} \{X_s > a - \epsilon\}.$$

  Com que aquest conjunt pertany a $\mathscr{F}_t$, T_a és un temps d'aturada.

  D'ara endavant, fem servir la notació $x \wedge y = \inf(x, y)$.

  Apliquem el teorema del Mostreig Opcional a la martingala $M_t = exp(\sigma X_t - (\sigma^2/2)t)$, amb $\sigma > 0$. Notem que no podem aplicar el teorema a T_a, ja que podria no ser acotat. Tot i això, si n és un enter positiu, llavors $T_a \wedge n$ és un temps d'aturada acotat (per l'últim punt de la Proposició 2.1.), i del teorema del Mostreig Opcional,

  $$\mathbb{E}(M_{T_a \wedge n}) = 1.$$

  On $M_{T_a \wedge n} = exp(\sigma X_{T_a \wedge n} - \sigma^2 (T_a \wedge n)/2) \leq exp(\sigma a)$.

  Per altra banda, si $T_a < +\infty$,

  $$\lim_{n \to +\infty} M_{T_a \wedge n} = M_{T_a}, \tag{2.1}$$

  i si $T_a = +\infty$, $X_t \leq a$ per tot t, i aleshores $\lim_{n \to +\infty} M_{T_a \wedge n} = 0$.

  Pel teorema de Lebesgue:

  $$\mathbb{E}\left(\mathbf{1}_{\{T_a < +\infty\}} M_{T_a}\right) = 1.$$

  Llavors, com $X_{T_a} = a$ quan $T_a < +\infty$,

  $$\mathbb{E}\left(\mathbf{1}_{\{T_a < +\infty\}} exp\left(-\frac{\sigma^2}{2} T_a\right)\right) = e^{-\sigma a}$$

  Si fem convergir σ a 0, obtenim $\mathbb{P}(T_a < +\infty) = 1$ (que vol dir que el Moviment Brownià arriba al nivell a quasi segurament). A més,

  $$\mathbb{E}\left(exp\left(-\frac{\sigma^2}{2} T_a\right)\right) = e^{-\sigma a}.$$

- El cas $a < 0$ es resol fàcilment si tenim en compte que

  $$T_a = \inf\{s \geq 0, -X_s = -a\},$$

  on $(-X_t)_{t\geq 0}$ és un $(\mathscr{F}_t)$-Moviment Brownià ja que és un proces estocàstic continu amb mitjana 0 i variància t i amb increments estacionaris i independents.

□

Una altra utilitat del teorema del Mostreig Opcional és calcular esperances fent servir els *màxims actualitzats (running maximum)* d'una martingala. Sigui M_t una martingala de quadrat integrable, es pot veure que el moment de segon ordre de $\sup_{0\leq t\leq T}|M_t|$ és acotat. És l'anomenada *Desigualtat de Doob:*

Proposició 2.6. ***(Desigualtat de Doob)*** *Si* $(M_t)_{0\leq t\leq T}$ *és una martingala continua, tenim*

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|M_t|^2\right)\leq 4\mathbb{E}(|M_T|^2).$$

Aquesta proposició, de fet, és un cas particular del *Teorema de desigualtats maximals*:

Teorema 2.4. ***(Desigualtats maximals)*** *Sigui* $(\mathscr{F}_t)_{t\geq 0}$ *una filtració en l'espai de probabilitat* $(\Omega,\mathscr{F},\mathbb{P})$, $(M_t)_{t\geq 0}$ *una martingala contínua respecte de la filtració* $(\mathscr{F}_t)_{t\geq 0}$ *i sigui* $p>1$ *i* $T>0$. *Si* $\mathbb{E}(|M_T|^p)<+\infty$, *aleshores:*

$$\mathbb{E}\left(\sup_{0\leq t\leq T}|M_t|^p\right)\leq\left(\frac{p}{p-1}\right)^p\mathbb{E}(|M_T|^p).$$

Abans demostrar aquest teorema, veiem-ne un altre que ens serà útil per a la prova:

Teorema 2.5. ***(d'aturada de Doob).*** *Sigui* $(\mathscr{F}_t)_{t\geq 0}$ *una filtració definida a l'espai de probabilitats* $(\Omega,\mathscr{F},\mathbb{P})$ *i* $(M_t)_{t\geq 0}$ *un procés estocàstic adaptat a la filtració* $(\mathscr{F}_t)_{t\geq 0}$, *i que les seves trajectòries* X_t *són contínues per la dreta i acotades localment. Aleshores, són equivalents:*

1. $(M_t)_{t\geq 0}$ *és una martingala respecte de la filtració* $(\mathscr{F}_t)_{t\geq 0}$.
2. *Per tot temps d'aturada acotat q.s.* T *de la filtració* $(\mathscr{F}_t)_{t\geq 0}$ *tal que* $\mathbb{E}(|M_T|)<+\infty$, *tenim* $\mathbb{E}(M_T)=\mathbb{E}(M_0)$.

Demostració. (del teorema 2.4)

- Suposem que $(M_t)_{t\geq 0}$ és una martingala respecte de la filtració $(\mathscr{F}_t)_{t\geq 0}$, i que les seves trajectòries X_t són contínues per la dreta i acotades localment. Agafem ara T un temps d'aturada de la filtració $(\mathscr{F}_t)_{t\geq 0}$ que està acotada quasi segurament per $K>0$. Suposem que T pren valors en un conjunt finit: $0\leq t_1<\cdots<t_n\leq K$. Per les propietats de martingala, tenim:

$$\begin{aligned}\mathbb{E}(M_T)&=\mathbb{E}\left(\sum_{i=1}^n M_T\mathbf{1}_{\{T=t_i\}}\right)=\sum_{i=1}^n\mathbb{E}\left(M_{t_i}\mathbf{1}_{\{T=t_i\}}\right)\\&=\sum_{i=1}^n\mathbb{E}\left(M_{t_n}\mathbf{1}_{\{T=t_i\}}\right)=\mathbb{E}(M_{t_n})=\mathbb{E}(M_0)\end{aligned}$$

  I la primera implicació del teorema queda provada quan T pren valors en un conjunt finit. Si T té un nombre infinit de valors, aleshores fem una aproximació de T amb la següent successió de temps d'aturada:

$$\tau_n=\sum_{k=1}^{2^n}\frac{kK}{2^n}\mathbf{1}_{\left\{\frac{(k-1)K}{2^n}\leq T\leq\frac{kK}{2^n}\right\}}$$

que pren valors en un conjunt finit quan $n \to +\infty$, $\tau_n \to T$. Per acabar aquesta part de la demostració, només queda veure que $\lim_{n\to+\infty} \mathbb{E}(M_{\tau_n}) = \mathbb{E}(M_T)$.

Per a fer-ho, provem que la família $(M_{\tau_n})_{n\in\mathbb{N}}$ és uniformement integrable: Sigui $A \geq 0$, com que τ_n pren valors en un conjunt finit, fent servir la propietat de martingala i la *desigualtat de Jensen*, és fàcil veure:

$$\mathbb{E}\left(|M_K|\mathbf{1}_{\{M_{\tau_n}\geq A\}}\right) \geq \mathbb{E}\left(|M_{\tau_n}|\mathbf{1}_{\{M_{\tau_n}\geq A\}}\right).$$

i llavors tenim:

$$\mathbb{E}(M_{\tau_n}\mathbf{1}_{\{M_{\tau_n}\geq A\}}) \leq \mathbb{E}(M_K\mathbf{1}_{\{\sup_{0\leq a\leq K} M_a\geq A\}}) \to_{A\to+\infty} 0.$$

Per integrabilitat uniforme i convergència en probabilitat, deduïm que:

$$\lim_{n\to+\infty} \mathbb{E}(M_{\tau_n}) = \mathbb{E}(M_T)$$

i aleshores podem concloure que $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

- Per altra banda, suposem ara que per tot temps d'aturada acotat T de la filtració $(\mathscr{F}_t)_{t\geq 0}$ tal que $\mathbb{E}(|M_T|) < +\infty$, tenim $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

  Siguin $0 \leq s \leq t$ i $A \in \mathscr{F}_s$, i fent servir el temps d'aturada $T = s\mathbf{1}_A + t\mathbf{1}_{A^c}$, obtenim:

  $$\mathbb{E}((M_t - M_s)\mathbf{1}_A) = 0,$$

  que implica la propietat de martingala per a $(M_t)_{t\geq 0}$.

□

Anem doncs amb la prova del teorema de Desigualtats maximals:

Demostració. (del teorema de Desigualtats maximals)
Sigui $p \geq 1$ i $T > 0$. $\mathbb{E}(|M_T|^p) < +\infty$ aleshores, per la *desigualtat de Jensen*, el procés $(|M_t|^p)_{0\leq t\leq T}$ és una submartingala. Sigui $\lambda > 0$ i $\tau = \inf\{s \geq 0 : |M_s| \geq \lambda\} \wedge T$, es veu que τ és un temps d'aturada acotat quasi segurament. Aleshores, pel *teorema d'aturada de Doob*:

$$\mathbb{E}(|M_\tau|^p) \leq \mathbb{E}(|M_T|^p).$$

Però de la definició de τ, tenim

$$|M_\tau|^p \geq \mathbf{1}_{\{\sup_{0\leq t\leq T}|M_t|\geq\lambda\}}\lambda^p + \mathbf{1}_{\{\sup_{0\leq t\leq T}|M_t|<\lambda\}}|M_T|^p.$$

que implica:

$$\mathbb{P}(\sup_{0\leq t\leq T}|M_t| \geq \lambda) \leq \frac{\mathbb{E}\left(|M_T|^p\mathbf{1}_{\{\sup_{0\leq t\leq T}|M_t|\geq\lambda\}}\right)}{\lambda^p} \leq \frac{\mathbb{E}(|M_T|^p)}{\lambda^p}.$$

que completa la prova de la primera part del teorema.

Siguin ara $p > 1$ i $T > 0$, suposem primer que $\mathbb{E}\left((\sup_{0\leq t\leq T}|M_t|)^p\right) < +\infty$, i tenim ja que per $\lambda > 0$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T}|M_t| \geq \lambda\right) \leq \frac{\mathbb{E}\left(|M_T|\mathbf{1}_{\{\sup_{0\leq t\leq T}|M_t|\geq\lambda\}}\right)}{\lambda},$$

llavors deduïm:

$$\int_0^{+\infty} \lambda^{p-1}\mathbb{P}\left(\sup_{0\le t\le T}|M_t|\ge\lambda\right)d\lambda \le \int_0^{+\infty}\lambda^{p-2}\mathbb{E}\left(|M_T|\mathbf{1}_{\{\sup_{0\le t\le T}|M_t|\ge\lambda\}}\right)d\lambda.$$

Del teorema de Fubini,

$$\begin{aligned}\int_0^{+\infty} \lambda^{p-1}\mathbb{P}\left(\sup_{0\le t\le T}|M_t|\ge\lambda\right)d\lambda &= \int_\Omega\left(\int_0^{\sup_{0\le t\le T}|M_t|(\omega)}\lambda^{p-1}d\lambda\right)d\mathbb{P}(\omega)\\ &= \frac{1}{p}\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right).\end{aligned}$$

i similarment, obtenim:

$$\int_0^{+\infty}\lambda^{p-2}\mathbb{E}\left(|M_T|\mathbf{1}_{\{\sup_{0\le t\le T}|M_t|\ge\lambda\}}\right)d\lambda = \frac{1}{p-1}\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^{p-1}|M_T|\right).$$

Aleshores, ajuntant els resultats,

$$\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)\le\frac{p}{p-1}\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^{p-1}|M_T|\right).$$

i usant ara la *desigualtat de Hölder* obtenim:

$$\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^{p-1}|M_T|\right)\le\mathbb{E}\left(|M_T|^p\right)^{\frac{1}{p}}\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)^{\frac{p-1}{p}}.$$

que implica que

$$\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)\le\frac{p}{p-1}\mathbb{E}\left(|M_t|^p\right)^{\frac{1}{p}}\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)^{\frac{p-1}{p}}.$$

En conclusió, si $\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)<+\infty$, tenim:

$$\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)\le\left(\frac{p}{p-1}\right)^p\mathbb{E}(|M_T|^p).$$

I per últim, si $\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_t|\right)^p\right)=+\infty$, considerem per a $N\in\mathbb{N}$ el temps d'aturada $\tau_N=\inf\{t\ge 0,|M_t|\ge N\}\wedge T$. Fent servir el resultat de dalt per a la martingala $(M_{t\wedge\tau_N})_{t\ge 0}$, obtenim:

$$\mathbb{E}\left(\left(\sup_{0\le t\le T}|M_{t\wedge\tau_N}|\right)^p\right)\le\left(\frac{p}{p-1}\right)^p\mathbb{E}(|M_T|^p),$$

i acabem la demostració fent servir el teorema de la convergència monòtona.

□

En un mercat viable i complet, es fa servir el teorema que es veurà a continuació per a associar qualsevol martingala amb estratègia d'inversió:

Teorema 2.6. ***(Descomposició de Doob)*** *Sigui* $(\Omega, \mathscr{F}, \mathbb{P})$ *un espai de probabilitat,* $I = \{1, 2, \cdots, N\}$ *un conjunt finit amb* $N \in \mathbb{N}$*,* $(\mathscr{F}_n)_{n\in I}$ *una filtració de* $\mathscr{F}$*, i* $X = (X_n)_{n\in I}$ *un proces estocàstic adaptat amb* $E|X_n| < \infty$ *per tot* $n \in I$*. Llavors existeix una martingala* $M = (M_n)_{n\in I}$ *amb un procés integrable* previsible $A = (A_n)_{n\in I}$*, començant per* A_0 *i tal que* $X_n = M_n + A_n$ *per tot* $n \in I$*. Aquí* previsible *vol dir que* A_n *és* $\mathscr{F}_{n-1}$*-mesurable per cada* $n \in I \setminus \{0\}$*. Aquesta descomposició és única quasi segurament.*

Demostració.
Per a l'existència, fem servir esperances condicionades. Definim els processos A i M com:

$$A_n = \sum_{k=1}^{n} (E[X_k|\mathscr{F}_{k-1}] - X_{k-1})$$

i

$$M_n = X_0 + \sum_{k=1}^{n} (X_k - E[X_k|\mathscr{F}_{k-1}])$$

on la suma per $n = 0$ està definida com a 0.

Veiem que A suma les esperances incrementals d'X, i M suma les inesperades, és a dir, la part de cada X_k que és desconeguda en el pas X_{k-1}.

Amb aquestes definicions, A_{n+1} i M_n són $\mathscr{F}_n$-mesurables ja que el procés X és adaptat, i $\mathbb{E}|A_n| < \infty$ i $\mathbb{E}|M_n| < \infty$ perquè el procés X és integrable, i la descomposició $X_n = M_n + A_n$ és vàlida per cada $n \in I$.

Per a la unicitat, prenem una altra descomposició $X = M' + A'$. Llavors el procés $Y := M - M' = A' - A$ és una martingala, i llavors:

$$E[Y_n|\mathscr{F}_{n-1}] = Y_{n-1} \; q.s.$$

i també és previsible, llavors:

$$E[Y_n|\mathscr{F}_{n-1}] = Y_n \; q.s.$$

per tot $n \in I \setminus \{0\}$. Com que per conveni $Y_0 = A'_0 - A_0 = 0$, això implica que $Y_n = 0$ q.s. per tot $n \in I$ i per tant la descomposició és única. $\square$

Capítol 3

Models de risc de crèdit

El mercat del crèdit i dels riscos derivats s'ha desenvolupat de manera exponencial en els últims anys. En aquest capítol, es fa una introducció als conceptes de modelització de risc de crèdit.

En la primera secció, es presenten els anomenats *Models estructurals*, en els que el temps d'incompliment està definit com un temps d'aturada respecte de la filtració associada al valor de la firma.

En la segona secció, s'introdueixen els *Models d'intensitat* que prenen el temps d'incompliment com a un temps aleatori extern, que ve caracteritzat per la *taxa de perillositat (Hazard rate)* . També es descriu la valoració dels CDS (*Credit Default Swaps* o pòlisses d'assegurança creditícia, en català).

En l'última secció es parla del concepte de *copula*, que es molt útil en models que tenen en compte varis temps d'incompliment.

Durant aquest capítol es va fent menció de conceptes explicats anteriorment al treball.

3.1 Models estructurals

Els *models estructurals* o *models del valor de la firma* proposen modelitzar un esdeveniment determinat relacionant-lo amb el valor de la firma, que teòricament seria el valor que s'hauria de pagar per fer-se amb l'empresa. Ens limitarem al model de Merton (*Robert C. Merton, 1944*), que és el pioner en aquesta branca.

En aquest model, el valor V_t de la firma a temps t segueix un *Moviment Brownià Geomètric*. El deute de la companyia està modelat per un *bonus cupó zero amb risc d'incompliment*, amb *maduresa* T i *valor nominal* $L > 0$. Més concretament, assumim que:

$$dV_t = V_t((r - k)dt + \sigma dW_t), \qquad V_0 > 0 \tag{3.1}$$

on $(W_t)_{t\geq 0}$ és un *Moviment Brownià Estàndard* sota la mesura de la probabilitat neutral al risc, $\mathbb{P}$. La constant r és el tipus d'interès i la constant k és la taxa de la despesa, i la *volatilitat* $\sigma > 0$ també és constant.

Amb aquest model, la companyia es declara en fallida si, a temps T, el valor de la firma és menor que el deute. En aquest cas, els deutors prenen control de la companyia. En altres paraules, els propietaris del bonus (deutors) reben, al *temps de maduresa* T, $V_T \wedge L$.

El deute de la companyia a temps $t < T$ ve donat per:

$$D_t = \mathbb{E}(e^{-r(T-t)}V_t \wedge L|\mathscr{F}_t)$$

on $\mathbb{F} = \{\mathscr{F}_t, t \geq 0\}$ és la *filtració natural* del procés V_t.

Notem que ja que $V_T \wedge L = L - (L - V_T)_+$, llavors tenim:

$$D_t = Le^{-r(T-t)} - \mathbb{E}(e^{-r(T-t)}(L - V_T)_+|\mathscr{F}_t)$$

així que el càlcul de D_t és equivalent al càlcul del preu d'una opció de venda europea del valor de la firma.

En aquesta situació, el valor d'equitat de la firma a temps T és la diferència entre el valor de la firma i el valor del deute:

$$E_T = V_T - V_T \wedge L = (V_T - L)_+$$

així que els accionistes de la firma poden veure's com els propietaris d'una opció de compra europea del valor de la firma de la companyia.

3.2 Models d'intensitat

Com hem comentat a la introducció del capítol, els models d'intensitat prenen el temps d'incompliment com a un temps aleatori extern, que ve caracteritzat per la *taxa de perillositat.*

3.2.1 Taxa de perillositat d'un temps aleatori

Proposició 3.1. *Sigui τ una variable aleatòria real, amb $\mathbb{P}(\tau > 0) = 1$ i $\mathbb{P}(\tau > t) > 0, \forall t > 0$. Si la funció de distribució de probabilitat de τ és de classe $\mathscr{C}^1$, llavors existeix una única funció λ que és no negativa i contínua i tal que:*

$$\forall t > 0, \quad \mathbb{P}(\tau > t) = e^{-\int_0^t \lambda(s)ds}$$

La funció λ es diu taxa de perillositat d'un temps aleatori τ.

Demostració.
Anomenem F a la funció de distribució de τ. Tenim $\mathbb{P}(\tau > t) = 1 - F(t)$. La funció $t \mapsto ln(1 - F(t))$ és contínuament diferenciable i no creixent, i la funció λ satisfà:

$$\lambda(t) = -\frac{d}{dt}ln(1 - F(t)) = \frac{F^{'}(t)}{1 - F(t)}, \quad t \geq 0.$$

Amb aquesta definició de λ ens trobem, fent servir que $\mathbb{P}(\tau > 0) = 1$;

$$\mathbb{P}(\tau > t) = e^{-ln(1-F(t))} = e^{-\int_0^t \lambda(s)ds}$$

□

Notem que per a que τ sigui un temps finit, la taxa de perillositat ha de satisfer $\int_0^\infty \lambda(t)dt = +\infty$. Notem també que tenim, per a $\delta > 0$;

$$\mathbb{P}(\tau \leq t + \delta|\tau > t) = \frac{\mathbb{P}(t < \tau \leq t + \delta)}{\mathbb{P}(\tau > t)} = 1 - e^{-\int_t^{t+\delta} \lambda(s)ds},$$

tal que

$$\lambda(t) = \lim_{\delta \to 0} \frac{\mathbb{P}(\tau > t + \delta | \tau > t)}{\delta}$$

Si τ modelitza un temps d'incompliment, la seva taxa de perillositat a temps t pot veure's com la taxa d'ocurrència d'incompliment al temps just després de t, donat que l'incompliment no ha ocorregut abans del temps t.

Remarca 3.1. *Un temps aleatori τ té una distribució exponencial si, i només si, la seva taxa de perillositat és constant. Per altra banda, si ξ és una variable aleatòria exponencial amb paràmetre 1, i si λ és una funció no negativa i contínua en $[0, +\infty)$, llavors la variable aleatòria*

$$\tau := \inf\{t \geq 0 | \int_0^t \lambda(s) ds \geq \xi\} \tag{3.2}$$

satisfà:

$$\mathbb{P}(\tau \leq t) = \mathbb{P}\left(\int_0^t \lambda(s) ds \geq \xi\right) = 1 - e^{\int_0^t \lambda(s) ds}$$

tal que la taxa de perillositat de τ és λ, i amb (3.2) construïm un temps aleatori donada una taxa de perillositat.

3.2.2 Intensitat i bonus cupó zero amb risc d'incompliment

Sigui $(\Omega, \mathscr{F}, \mathbb{P})$ un espai de probabilitat. La informació relacionada amb el mercat sense incompliments es denota com $(\mathscr{F}_t)_t \geq 0$. Normalment aquesta filtració té incorporat l'historial de taxes d'interès, tal que si denotem per (S_t^0) el proces de preus de l'actiu sense risc, i per $(r(t))$ el procés amb taxa d'interès instantani, llavors aquests dos processos són $(\mathscr{F}_t)$-adaptats.

El temps d'incompliment es modelitza amb una variable aleatòria τ, que pot *no* ser un temps d'aturada respecte de la filtració sense incompliments, $(\mathscr{F}_t)$. Aquesta és una diferència amb els models estructurals. En els models d'intensitat, el temps d'incompliment és una variable exògena i l'incompliment pot ocórrer per motius al·liens al model, de manera inesperada.

Per l'altra banda, en un temps donat t, els inversors saben si s'ha donat l'incompliment o no. Aleshores, la informació total disponible a temps t es descriu com el conjunt $\mathbb{G}$, generat per $(\mathscr{F}_t)$ i per $\{\tau \leq s\}$, per $s \in [0, t]$:

$$\mathscr{G}_t = \mathscr{F}_t \vee \sigma(\tau \leq s, s \leq t).$$

Observacions. *τ és un temps d'aturada respecte de la filtració $(\mathscr{G}_t)$. Una* mesura de preus *neutral al risc, en aquest context, és una mesura de probabilitat $\mathbb{P}^* \sim \mathbb{P}$, sota la qual els preus descomptats dels actius amb risc són $(\mathscr{G}_t)$-martingales.*

Els preus dels actius amb risc que poden ser afectats per l'incompliment són adaptats *a la filtració $(\mathscr{G}_t)$.*

Es considera ara un *bonus cupó zero amb risc d'incompliment*, amb maduresa T, que a temps T paga una unitat monetària si no hi ha incompliment, i no paga res si hi ha hagut incompliment abans o en el temps T. El saldo d'aquest derivat a temps T ve donat per

$\mathbf{1}_{\{\tau>T\}}$. Llavors el seu valor a temps t ve donat per l'esperança condicionada, donada $\mathscr{G}_t$, respecte de la mesura de preus del saldo descomptat, i.e.

$$D(t,T) = E^*\left(\mathbf{1}_{\{\tau>T\}} e^{-\int_t^t r(s)ds} | \mathscr{G}_t\right), \quad 0 \le t \le T.$$

La proposició següent relaciona el càlcul de l'esperança condicionada donada $\mathscr{G}_t$ amb l'esperança condicionada donada $\mathscr{F}_t$:

Proposició 3.2. *Per tota variable aleatòria X no negativa, amb probabilitat 1 es té:*

$$\mathbf{1}_{\{\tau>T\}} E^*(X|\mathscr{G}_t) = \mathbf{1}_{\{\tau>T\}} \frac{E^*(X\mathbf{1}_{\{\tau>T\}}|\mathscr{F}_t)}{E^*(\mathbf{1}_{\{\tau>T\}}|\mathscr{F}_t)}$$

Demostració. Observem que, al conjunt $\{\tau > t\}$, tenim $E^*\left(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right) > 0$ quasi segurament, així que la variable aleatòria:

$$Y := \mathbf{1}_{\{\tau>t\}} \frac{E^*\left(X\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}{E^*\left(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}$$

està ben definida i és $\mathscr{G}_t$-mesurable .

Per altra banda, tenim

$$\mathbf{1}_{\{\tau>t\}} E^*(X|\mathscr{G}_t) = E^*\left(X\mathbf{1}_{\{\tau>t\}}|\mathscr{G}_t\right),$$

ja que $\{\tau > t\} \in \mathscr{G}_t$. Per a provar que $Y = E^*\left(X\mathbf{1}_{\{\tau>t\}}|\mathscr{G}_t\right)$, només cal veure que

$$\forall A \in \mathscr{C}, \quad E^*\left(X\mathbf{1}_{\{\tau>t\}}\mathbf{1}_A\right) = E^*(Y\mathbf{1}_A),$$

on $\mathscr{C}$ és una subclasse de $\mathscr{G}_t$, estable sota interseccions finites, i genera $\mathscr{G}_t$ (això es pot veure, per exemple, a *Jacod, J; Protter P. (2003): Probability Essentials. Springer*).

Sigui, doncs, $\mathscr{C}$ una classe d'esdeveniments A que es poden expressar de la forma:

$$A = \{r \le s\} \cap B,$$

on $B \in \mathscr{F}_t$ i $s \in [0,t] \cup \{+\infty\}$. La classe $\mathscr{C}$ és estable sota interseccions finites i genera $\mathscr{G}_t$. Per tota $s \in [0,t]$, tenim $E^*\left(X\mathbf{1}_{\{\tau>t\}}\mathbf{1}_{\{\tau\le s\}\cap B}\right) = 0$ i $E^*\left(Y\mathbf{1}_{\{\tau\le s\}\cap B}\right) = 0$. Per $s = \infty$, tenim $A = B$ i, com que $B \in \mathscr{F}_t$,

$$E^*\left(X\mathbf{1}_{\{\tau>t\}}\mathbf{1}_B\right) = E^*(E^*(X\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t)\mathbf{1}_B).$$

També tenim:

$$\begin{aligned} E^*(Y\mathbf{1}_B) &= E^*\left(\mathbf{1}_{\{\tau>t\}} \frac{E^*\left(X\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}{E^*\left(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}\mathbf{1}_B\right) \\ &= E^*\left[E^*(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t)\frac{E^*\left(X\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}{E^*\left(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}\mathbf{1}_B\right] \\ &= E^*(X\mathbf{1}_{\{\tau>t\}}\mathbf{1}_B). \end{aligned}$$

$\square$

Ara, suposem coneguda la taxa de perillositat condicionada de τ, donada la filtració sense incompliments. Més concretament, se suposa que existeix un procés $(\mathscr{F}_t)$-adaptat, no negatiu, $(\lambda(t))_{t\geq 0}$ tal que, per tot $t \geq 0$,

$$\mathbb{P}^*(\tau > t|\mathscr{F}_t) = E^*(\mathbf{1}_{\{\tau>T\}}|\mathscr{F}_t) = e^{-\int_0^t \lambda(s)ds}.$$

Usant la proposició 3.2, amb $X = \mathbf{1}_{\{\tau>T\}}e^{\int_t^T r(s)ds}$, obtenim el valor per al *bonus cupó zero amb risc d'incompliment* abans de l'incompliment:

$$\begin{aligned}\mathbf{1}_{\{\tau>T\}}D(t,T) &= \mathbf{1}_{\{\tau>T\}}\frac{E^*\left(\mathbf{1}_{\{\tau>T\}}e^{-\int_t^T r(s)ds}\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)}{E^*\left(\mathbf{1}_{\{\tau>t\}}|\mathscr{F}_t\right)} \\ &= \mathbf{1}_{\{\tau>T\}}\frac{E^*\left(\mathbf{1}_{\{\tau>T\}}e^{-\int_t^T r(s)ds}|\mathscr{F}_t\right)}{e^{-\int_0^t \lambda(s)ds}} \\ &= \mathbf{1}_{\{\tau>T\}}\frac{E^*\left(e^{-\int_0^T \lambda(s)ds}e^{-\int_t^T r(s)ds}|\mathscr{F}_t\right)}{e^{-\int_0^t \lambda(s)ds}}\end{aligned}$$

on l'última igualtat es segueix de condicionar respecte la $\mathscr{F}_T$ i fer servir la $\mathscr{F}_T$-mesurabilitat de $e^{-\int_t^T r(s)ds}$. Llavors, també hem provat el resultat següent:

Proposició 3.3. *El valor a temps t d'un* bonus cupó zero amb risc d'incompliment*, amb maduresa T abans d'incompliment ve donat per:*

$$\tilde{D}(t,T) = E^*\left(e^{-\int_t^T (r(s)+\lambda(s))ds}|\mathscr{F}_t\right), \quad 0 \leq t \leq T,$$

on $(\lambda(t))$ es la taxa de perillositat condicional d'un incompliment donada la filtració sense defaults.

El procés $(\lambda(t))$ també s'anomena *intensitat* de l'incompliment.

3.2.3 Credit Default Swaps (CDS)

Un CDS o *Credit Default Swap* és un derivat que ofereix protecció d'incompliments. Un inversor A (l'anomenarem *Branca Premium*), que vol estar protegit d'incompliments, arriba a l'acord següent amb un banc B(l'anomenarem *Branca Protecció*): A dates $T_1, \cdots, T_n$, A realitzarà uns pagaments fixos a B, sempre que no hi hagi incompliment. Per altra banda, en cas d'incompliment abans del temps T_n, B farà un pagament a A. Els *cashflows* (terme que descriu els fluxos de capital) pagats per A i per B estan descrits, respectivament, per la *branca premium* i per la *branca protecció*, així:

- *Branca Premium*: cashflow $sN(T_i - T_{i-1})\mathbf{1}_{\{\tau>T_i\}}$ a temps $T_i, i = 1, \cdots, T_n, (T_0 = 0)$. El nombre N es diu valor *nominal* del CDS, i s és la *propagació* del CDS.
- *Branca Protecció*: cashflow $N(1-R)\mathbf{1}_{\{\tau\leq T_n\}}$ a temps τ, on R és la *taxa de recuperació*. La idea és que A té un bonus amb nominal N emès per una empresa C que pot caure en incompliment, i que en aquest cas, el propietari del bonus recupera *NR*, enlloc de recuperar *N*. Així doncs, el pagament de protecció ha de ser *N(1-R)*. La taxa de recuperació real no és coneguda, el conveni és prendre $R = 40\%$.

Remarca 3.2. *Com que els* Credit Default Swaps *són els derivats crediticis més líquids, normalment es fan servir per a calibrats de paràmetres ("recalcular"paràmetres afegint canvis al model) com ara amb la propagació dels* CDS *amb vàries madureses, que es fan servir per calibrar probabilitats d'incompliment..*

3.3 Còpules

Una *còpula* és una funció de distribució conjunta d'un vector de variables aleatòries uniformes en [0,1].

Definició 3.1. *Una funció $C : [0,1]^m \to [0,1]$ es diu còpula si existeix un vector aleatori $(U_1, \cdots, U_m)$ tal que cada U_i està distribuïda uniformement en [0,1] i*

$$C(u_1, \cdots, u_m) = \mathbb{P}(U_1 \leq u_1, \cdots, U_m \leq u_m), \quad (u_1, \cdots, u_m) \in [0,1]^m.$$

El resultat següent, conegut com *Teorema d'Sklar*, mostra que la llei d'un vector aleatori queda caracteritzara amb termes de la seva distribució marginal i una còpula:

Teorema 3.1. *Sigui $X = (X_1, \cdots, X_m)$ un vector aleatori amb valors en $\mathbb{R}^m$. Per $i = 1, \cdots, m$, denotem per F_i la funció de distribució de la variable aleatòria X_i, $(F_i(x) = \mathbb{P}(X_i \leq x))$. Aleshores, existeix una còpula C tal que, per tot $(x_1, \cdots, x_m) \in \mathbb{R}^m$,*

$$\mathbb{P}(X_1 \leq x_1, \cdots, X_m \leq x_m) = C(F_1(x), \cdots, F_m(X_m)).$$

A més, si les funcions F_i són contínues, la còpula C és única.

Abans de veure la demostració d'aquest teorema, veiem el lema següent:

Lema 3.1. *Si X és una variable aleatòria de valors reals, amb una funció de distribució F contínua, llavors la variable aleatòria F(X) està distribuïda uniformement en [0,1].*

Demostració. Per $u \in (0,1)$, sigui

$$G(u) = \inf\{x \in \mathbb{R} | F(x) \geq u\}.$$

Notem que G(u) està ben definida ja que

$$\lim_{x \to -\infty} F(x) = 0 \quad i \quad \lim_{x \to \infty} F(x) = 1.$$

Com que F és contínua per la dreta, tenim que $F(G(u)) \geq u$. De fet, com que F és contínua, tenim que $F(G(u)) = u$ (si tinguéssim $F(G(u)) > u$, aleshores $F(x) > u$ per a x propera a *G(u)* i menor que *G(u)*, que seria una contradicció amb la definició de *G(u)*). Per altra banda, és fàcil provar que $F(x) \geq u$ si i només si $x \geq G(u)$. Per tant, tenim, fent servir $\mathbb{P}(X = G(u)) = 0$,

$$\begin{aligned}
\mathbb{P}(F(X) \geq u) &= \mathbb{P}(X \geq G(u)) \\
&= \mathbb{P}(X > G(u)) \\
&= 1 - \mathbb{P}(X \leq G(u)) \\
&= 1 - F(G(u)) \\
&= 1 - u,
\end{aligned}$$

Es veu, per tant, que *F(X)* és uniforme en [0,1], i demostrem així el lema.

□

Anem ara a veure la demostració del *Teorema d'Sklar*.

Demostració. (Teorema d'Sklar). Provarem el teorema per a marginals contínues. Sabem, pel Lema anterior, que les variables aleatòries $F_i(X_i)$ estan uniformement distribuïdes en [0,1]. Sigui C la còpula:

$$C(u_1, \cdots, u_m) = \mathbb{P}(F_1(X_1) \leq u_1, \cdots, F_m(X_m) \leq u_m), \quad (u_1, \cdots, u_m) \in [0,1]^m.$$

Com que F_i és no decreixent, tenim $\{X_i \leq x_i\} \subset \{F_i(X_i) \leq F_i(x_i)\}$. A més,

$$\mathbb{P}(X_i \leq x_i) = F_i(x_i) = \mathbb{P}(F_i(X_i) \leq F_i(x_i)),$$

on l'última desigualtat es segueix del mateix Lema 3.1. Aleshores, els esdeveniments $\{X_i \leq x_i\}$ i $\{F_i(X_i) \leq F_i(x_i)\}$ són els mateixos quasi segurament, i llavors tenim:

$$\begin{aligned}\mathbb{P}(X_1 \leq x_1, \cdots, X_m \leq x_m) &= \mathbb{P}(F_1(X_1) \leq F_1(x_1), \cdots, F_m(X_m) \leq F_m(x_m)) \\ &= C(F_1(x_1), \cdots, F_m(x_m)).\end{aligned}$$

Veiem aquí que la unicitat de C es segueix del fet que qualsevol vector $(u_1, \cdots, u_m)$ amb $0 < u_i < 1$ poden escriure's com $(F_1(x_1), \cdots, F_m(x_m))$ per alguns $x_i's$, per la continuïtat de les $F_i's$. □

Es veuran ara unes propietats de les còpules:

Propietats.

- *Si $(X_1, \cdots, X_n)$ té una còpula C i si $f_1, \cdots, f_m$ són funcions estrictament creixents i contínues, aleshores el vector $(f_1(X_1), \cdots, f_m(X_m))$ té la mateixa còpula C.*
- *Les variables aleatòries $X_1, \cdots, X_m$ són independents si i només si el vector $(X_1, \cdots, X_m)$ admet la còpula $C(u_1, \cdots, u_m) = \prod_{i=1}^m u_i$.*
- *Si $X_1 = \cdots = X_m$, aleshores el vector $(X_1, \cdots, X_m)$ té una còpula $C(u_1, \cdots, u_m) = \min_{1 \leq i \leq m} u_i$.*
- *Cotes Fréchet-Hoeffding: Per tota còpula C, tenim*

  $$C^-(u_1, \cdots, u_m) \leq C(u_1, \cdots, u_m) \leq C^+(u_1, \cdots, u_m),$$

  on $C^-(u_1, \cdots, u_m) = \left(\sum_{i=1}^m u_i - (m-1)\right)_+$ i $C^+(u_1, \cdots, u_m) = \min_{1 \leq i \leq m} u_i$. Observem que, per a $m \geq 3$, la funció C^- no és una còpula.

Remarca 3.3. *Les còpules es fan servir per a quantificar derivats d'instruments financers crediticis que tinguin incompliments de diverses empreses. Els models requereixen la definició de diferents temps d'incompliment, $\tau_1, \cdots, \tau_m$, i a aquests models se'ls hi posen diferents noms. El* primer contracte en incomplir*, al que típicament se li involucra la distribució de* $\min(\tau_1, \cdots, \tau_m)$*. Les distribucions marginals de $\tau_1, \cdots, \tau_m$ són conegudes (fent calibrats de* Credit Default Swaps *en els models), i es fa servir una còpula per a produir la distribució conjunta, comunament anomenada* CDO (Collateralized Debt Obligations).

A la pràctica, és important saber parametritzar famílies de còpules ajustant-les a les dades dels mercats financers. Per exemple, una família de còpules importants en finances és la família de Còpules Gaussianes *(és a dir, còpules de vectors Gaussians). Poden ser parametritzades per la matriu de covariàncies del vector Gaussià.*

Capítol 4

Machine Learning al camp del Risc de Crèdit

Els mètodes actualment emprats pels bancs i entitats financeres per a la predicció i previsió d'impagaments són els més senzills d'entendre i d'implementar: els models lineals generalitzats. Concretament, els models de regressions (lineals, logístiques, exponencials...) o també models que inclouen un factor de penalització, com ara el model logístic amb penalització LASSO (*Least Absolute Shrinkage and Selection Operator*). Veiem-ne unes definicions per a entendre els seus avantatges i inconvenients.

4.1 Regressions

L'anàlisi regressiu és una metodologia estadística que permet determinar el nivell de relació entre dues o més variables. Els resultats d'una regressió ajuden a predir un valor desconegut estudiant-ne la relació entre les variables. Per exemple, entre l'altura i el pes d'una persona hi ha relació: generalment, les persones més altes tendeixen a pesar més. Aleshores, podríem predir el pes d'un individu, donada l'altura. A les variables que es fan servir com a base predictiva del model se'ls hi diuen *independents o explicatives*, i a la variable que s'estima o es prediu se l'anomena *dependent o resposta*.

En cas de que no es pugui aplicar un model de regressió a un estudi, es diu que no hi ha *correlació* entre les variables estudiades. El model es pot expressar com:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \epsilon$$

on Y és la variable *dependent o resposta*, $X_1, \cdots, X_m$ són les variables *independents o explicatives*, $\beta_0, \beta_1, \cdots, \beta_m$ són els paràmetres del model, que mesuren la influència de les variables explicatives sobre les de resposta, i ϵ és *l'error d'ajustament*.

4.1.1 Regressió lineal

Regressió lineal simple

Quan el model només està conformat per dues variables estadístiques, X i Y, on X és la variable *independent o explicativa*, i Y és la variable *dependent o resposta*, a la regressió se l'anomena *regressió lineal simple*. És relacionen mitjançant la relació funcional:

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

on $\beta_0, \beta_1 \in \mathbb{R}$ són constants desconegudes que es diuen coeficients de regressió, i que s'han d'estimar, per exemple, pel mètode dels mínims quadrats, que s'explica més endavant.

Regressió lineal múltiple

Si hi ha més variables independents, es diu *regressió lineal múltiple*. Aquest anàlisi de correlació es fa a través d'equacions amb diferents variables independents, i diferents paràmetres, amb la relació funcional:

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \epsilon = \beta_0 + \sum_{j=1}^{m} \beta_j X_j + \epsilon,$$

on els $\beta_j, \forall j \in \{0, 1, \cdots, m\}$ són els anomenats coeficients del model de regressió múltiple, i per a estimar-los s'utilitza sovint el mètode dels mínims quadrats.

Mètode dels mínims quadrats

Suposem que tenim un model de regressió lineal múltiple, i que tenim la funció d'error quadràtic que volem minimitzar:

$$S(\beta_0, \beta_1, \cdots, \beta_m) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{m} \beta_j x_{ij} \right)^2,$$

on ϵ_i és l'error associat a la i-èssima mesura del valor X_j (i s'acostuma a fer servir que $\epsilon_i \sim N(0, \sigma^2)$), y_i denota la i-èssima observació de Y i x_{ij} denota l'i-èssim valor observat de X_j.

Els *estimadors per mínims quadrats* denotats per $\hat{\beta}_0, \cdots, \hat{\beta}_m$ han de satisfer:

$$\left. \frac{\partial S}{\partial \beta_j} \right|_{\hat{\beta}_0, \cdots, \hat{\beta}_m} = 0$$

per tot $j \in \{0, \cdots, m\}$.

Davant de la dificultat de resoldre aquest sistema de m+1 equacions de forma analítica, sovint s'escriu el model de regressió lineal múltiple

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_m X_m + \epsilon$$

en forma matricial com:

$$Y = X\beta + \epsilon,$$

amb:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

En forma matricial, la funció d'error quadràtic pot expressar-se com:

$$S(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon^T \epsilon = (Y - X\beta)^T (Y - X\beta)$$

i la obtenció dels estimadors es realitza resolent el sistema lineal d'equacions.

Els estimadors per mínims quadrats han de satisfer:

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = \mathbf{0}$$

on $\hat{\beta} \in \mathbb{R}^{m+1x1}$ denota el vector que conté als estimadors, i $\mathbf{0}$ és un vector de zeros.

Llavors la condició anterior es redueix a:

$$X^T X \hat{\beta} = X^T Y$$

Si la matriu inversa $(X^T X)^{-1}$ existeix, aleshores l'estimador per mínims quadrats ve donat per:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Pel que el model ajustat de regressió múltiple ve donar per:

$$\hat{y} = X^T \hat{\beta} = \hat{\beta}_0 + \sum_{j=1}^{m} \hat{\beta}_j x_j = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_m x_m$$

4.1.2 Regressió logística

La regressió logística és un altre tipus d'anàlisi de regressió utilitzar per a predir el resultat d'una variable categòrica (que pot adoptar un nombre limitat de categories) en funció de les variables independents predictores. És utilitzada sovint en ciències mèdiques i socials.

Aquesta regressió analitza dades distribuïdes binomialment de la forma:

$$Y_i \sim B(p_i, n_i),\ per\ i = 1, \cdots, m,$$

on els nombres d'assaigs Bernoulli n_i són coneguts i les probabilitats d'èxit p_i són desconegudes. Per exemple, aquest és el cas de la distribució percentual de llavors (p_i) que germinen després de plantar-ne (n_i),

El model s'obté a base del que cada assaig (cada valor d'i) i el conjunt de variables explicatives poden informar sobre la probabilitat final. Aquestes variables poden pensar-se com un vector X_i, i el model pren llavors la forma:

$$p_i = E\left(\frac{Y_i}{n_i} \middle| X_i \right)$$

4.1.3 Regressió exponencial

En determinats experiments, majoritàriament biològics, la dependència entre les variables X i Y és de forma exponencial, de la forma:

$$Y = ae^{bX}$$

Mitjançant una transformació lineal, prenent logaritmes, es pot convertir en un problema de regressió lineal com el que hem explicat anteriorment:

$$\ln(Y) = bX + \ln(a)$$

Hi ha molts més tipus de regressions, però passem ara a explicar el tipus de models més moderns, on hi intervenen algoritmes de Machine Learning.

4.2 Models de Machine Learning

En un món que està en desenvolupament constant, sobretot tecnològicament parlant, està clar que aquests mètodes tradicionals també es veuran actualitzats d'una manera o una altra.

Des dels anys 50 que existeixen coneixements matemàtics per a l'ús de metodologies derivades de la tant aclamada intel·ligència artificial, que inclou també el Machine Learning, però la seva implementació i utilització només ha estat possible en aquests darrers lustres. A més, gairebé tots els experts en la matèria coincideixen que tot just s'acaba de descobrir, i que cada any se n'aprenen *infinitat* de coses noves.

El Machine Learning es podria definir com un conjunt de tècniques estadístiques que permeten als ordinadors aprendre i millorar únicament a partir d'introduir-hi inicialment una gran quantitat de dades. És clar que el futur dels models estadístics passa per l'aprenentatge automàtic dels programes i que cal invertir-hi els recursos que siguin necessaris.

Concretament en el cas del risc creditici, té molt sentit implementar eines derivades del Machine Learning ja que poden usar-se per a la predicció i classificació de clients segons la puntuació, simplement introduint les dades històriques dels préstecs que hagi concedit un banc en un cert període de temps. De fet, segons l'Institut Financer Internacional, un 37% d'una mostra significativa d'empreses a nivell global utilitzen ja un procés íntegre amb aquestes tècniques en el sector del crèdit *Scoring*.

I es que és cert i conegut que els algoritmes de Machine Learning són capaços de millorar considerablement la capacitat predictiva en la majoria de situacions respecte dels models tradicionals emprats fins ara, i si la implementació completa fos d'un dia per l'altre, tothom s'hi passaria sense pensar-s'ho dues vegades. Però precisament aquest és el problema: malgrat la millora evident, també pot produir molts mals de cap a causa de la seva complexitat tecnològica i dificultat cognitiva, fet que ha comportat les restriccions del Comitè de Supervisió Bancària de Basilea respecte a la utilització d'aquestes metodologies. Es troben davant del repte de permetre que es maximitzin les oportunitats derivades del progrés tecnològic i la innovació financera, alhora que es segueixi respectant la neutralitat tecnològica i la compatibilitat amb les regulacions ja establertes. Això li dóna sentit als resultats d'una enquesta que van fer ells mateixos, ja que mostren una davallada de l'interès en l'ús d'aquesta metodologia al passar d'un 60% l'any 2018 al 40% l'any 2020. Les empreses constaten que necessiten models simples i fàcils d'interpretar i implantar, mentre que els models de Machine Learning són més complexos d'entendre (tot i que lluny d'impossible), i tenint en compte que no hi ha els mateixos recursos tecnològics a tot arreu, ja que el desenvolupament en aquest camp que hi ha a Europa, a Amèrica del Nord o a Xina no és el mateix que a altres regions.

Per tots aquests motius és necessari que hi hagi unes guies per a entendre i arribar a dominar els nous algoritmes d'aprenentatge automàtic, alhora que unes normes que limitin les despeses d'inversions en noves tecnologies. Aquestes eines i mesures (i moltes més) venen recollides en el conjunt de regulacions anomenat Basilea III.

4.3 Quantificació dels beneficis

Una dificultat en la recerca d'aquest camp és que hi ha molt poca disponibilitat de bases de dades de llargues sèries temporals d'alta qualitat. Això és perquè acostumen a estar mal enregistrades perquè històricament en el sector bancari se li ha dona més valor al curt termini que al llarg, i no es pateix per deixar enregistrades transaccions que ja estan resoltes i comprovar la veracitat de les dades ja guardades.

Tot i així, és necessari establir fórmules per a mesurar els beneficis i els costos dels models, per a veure si realment val la pena implementar-los. Quan es fa referència als beneficis, sovint es parla de la capacitat de predicció, del nombre d'impagaments predits adequadament o del nombre de prediccions totals correctes.

La mètrica més utilitzada per a establir els guanys és l'anomenada *àrea sota la corba (AUC)*, ja que permet contenir amb un únic valor tota la informació que interessa per a analitzar el model. Una altra mètrica molt utilitzada per al risc de crèdit és l'índex de Gini, que quantifica la desigualtat i diferències entre grups heterogenis de clients, segons el risc creditici.

4.3.1 Corba ROC i àrea sota la corba (AUC)

Una *Corba ROC (Receiver o Relating Operating Characteristic)* és una representació gràfica de la *sensibilitat* davant de l'*especificitat* d'un sistema classificatori binari, com ara seria si un client està en *Default* o no. Aquest sistema també es pot interpretar com la representació de la proporció entre la taxa de vertaders positius i la de falsos positius, segons varia el llindar de discriminació (a partir de quan prenem el client com a *Default*), i per tant es pot també definir com una probabilitat entre 0 i 1.

La metodologia que utilitza la Corba ROC, o simplement el *Mètode ROC*, és fa servir molt per a observar el comportament dels models a l'hora de classificar individus. De fet, es fa servir un derivat: l'*àrea sota la corba (AUC)*, que com el seu nom indica, mesura l'àrea que queda sota la corba de probabilitat, i serveix per informar de la capacitat que té el model per distingir entre les dues possibles classes. Com més s'aproximi a 1, millor capacitat predictiva tindrà un model.

Per a representar-la, com s'ha explicat, prenem la taxa de vertaders positius (TPR), o *sensibilitat*, que és la taxa de positius que han estat ben detectats pel model, i la posem a l'eix de coordenades. A l'eix d'abcisses hi posem la taxa de falsos positius (FPR), o *ratio*, que es la taxa de casos negatius que el model ha detectat com a positius i la podem calcular com el complementari de l'*especificitat*, que és la taxa de vertaders negatius (negatius que han estat ben detectats pel model):

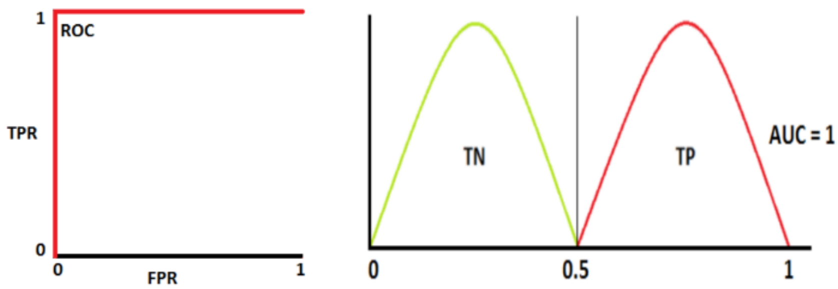
$$Sensibilitat\ (TPR) = \frac{Vertaders\ positius}{Vertaders\ positius + Falsos\ negatius}$$

$$Especificitat = \frac{Vertaders\ negatius}{Falsos\ positius + Verdaders\ negatius}$$

$$Ratio\ (FPR) = 1 - Especificitat$$

Podem trobar quatre tipus de corba ROC amb la seva corresponent AUC:

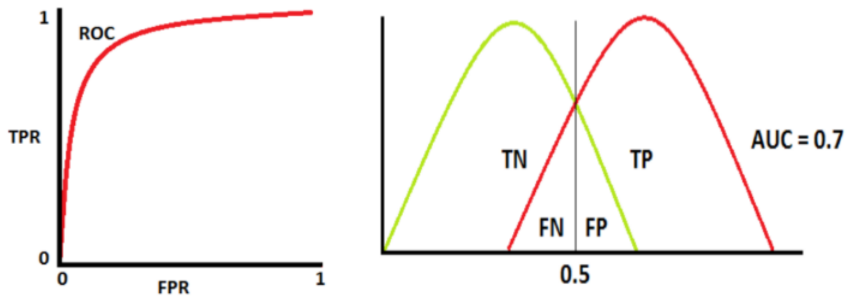
1. El model té una predicció perfecta, és a dir, que detecta tots els casos positius (per exemple, clients que estaran en *Default*) com a positius i tots els casos negatius (clients que no incompliran) com a negatius:



Figures 1 i 2. Gràfics ROC i AUC perfectes. (Narkhede, 2018)

La figura 1 es la corba ROC d'una situació ideal, gairebé impossible d'aconseguir a la pràctica, tot i que sí que es habitual que es trobin valors pròxims a 1. Veiem que l'àrea sota la corba és 1, ja que com que tot ha estat predit correctament, la taxa de vertaders positius és d'1 i la de falsos negatius és 0. En la figura 2 prenem, per exemple, 0.5 com a valor crític per a distingir si es pren com a positiu o com a negatiu, i veiem que la intersecció entre vertaders positius (TP) i vertaders negatius (TN), que serien els falsos positius i falsos negatius, és nul·la.

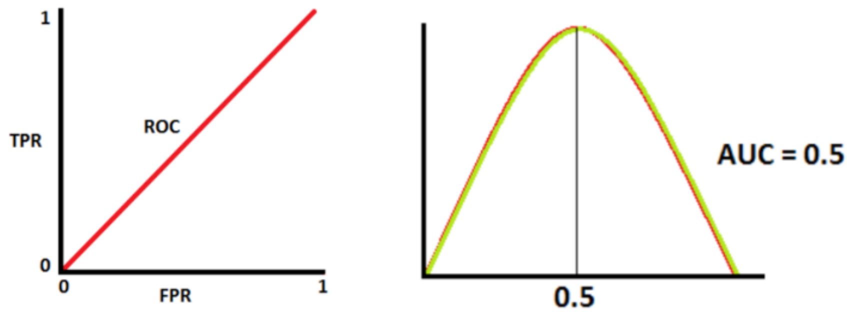
2. El model fa una bona predicció, és a dir, que detecta la majoria dels casos positius com a positius i la majoria dels casos negatius com a negatius, però té cert marge de millora:



Figures 3 i 4. Gràfics ROC i AUC bons. (Narkhede, 2018)

La figura 3 es la corba ROC d'una situació ben quotidiana, que és que s'aconsegueix que el model predigui força bé tots els casos, i l'objectiu és anar modificant-lo per a que cada vegada ho faci millor. L'AUC és superior a 0.5 però inferior a 1 (en aquest cas és de 0.7). El model no és perfecte (veiem que la intersecció és no buida), però és capaç de detectar certs patrons en les dades.

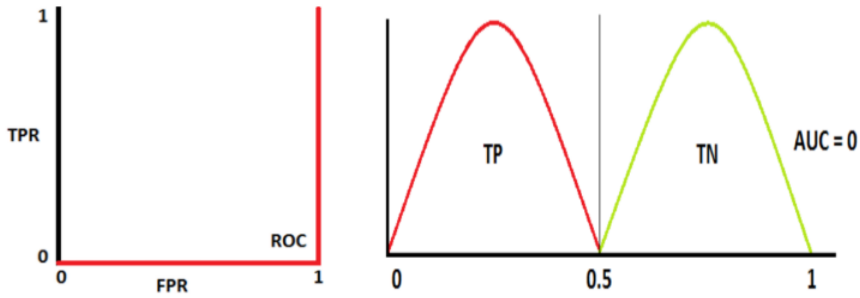
3. El model no sap classificar les observacions. Aquest seria el pitjor dels casos a l'hora d'avaluar un model, ja que significaria que no té capacitat per a discriminar entre casos positius (és a dir, prestataris que incompleixen) i casos negatius (és a dir, prestataris que no incompleixen):



Figures 5 i 6. Gràfics ROC i AUC dolents. (Narkhede, 2018)

La figura 5 es la corba ROC, que veiem que és la recta $y = x$, l'àrea sota de la corba seria de 0.5 i la intersecció entre avaluacions de vertaders i falsos negatius i positius és total.

4. Per últim tenim el model que fa una predicció completament errònia. Aquest seria el cas contrari al primer model, ja que tot i que també sigui un cas gairebé impossible a la pràctica, aquest prediu erròniament totes les observacions. En el cas del risc de crèdit, seria que es classifiquessin com a sans a tots els clients que acaben en *Default* i a l'inrevés. Normalment es deu a una equivocació humana a l'hora de programar el model:



Figures 7 i 8. Gràfics ROC i AUC contraris. (Narkhede, 2018)

La figura 7 es la corba ROC, l'àrea sota de la corba seria 0 i la intersecció entre positius i negatius és 0, però en aquest cas perquè tots els negatius han sigut avaluats com a positius i tots els positius com a negatius.

4.3.2 Índex de Gini

L'índex de Gini o coeficient de Gini (*Corrado Gini, 1884-1965*) és una mesura molt utilitzada en econometria. Normalment es fa servir per a mesurar la desigualtat d'ingressos dins d'un mateix país, però es pot fer servir per a mesurar qualsevol tipus de distribució, ja que en quantifica les seves desigualtats. El coeficient de Gini pren valors entre 0 i 1, i l'índex de Gini és el coeficient de Gini expressat percentualment, de manera que 0 representa la perfecta igualtat, i 1 (o 100) la perfecta desigualtat. Per exemple, 0 seria que totes les persones del país cobren el mateix, i 1 (o 100) seria que una sola persona cobri tots els ingressos del país, i la resta cap.

L'índex o coeficient de Gini es calcula com una proporció entre les àrees del diagrama de la *Corba de Lorenz (Max Otto Lorenz, 1876-1959)*, ja que aquest representa gràficament la distribució relativa d'una variable dins d'una població, essent cada un dels punts de la corba un percentatge acumulatiu dels possibles valors del domini, i l'índex de Gini és la relació de l'àrea que compresa entre la corba i la recta $y = x$ entre l'àrea de sota la recta $y = x$.

Veiem-ne un exemple:

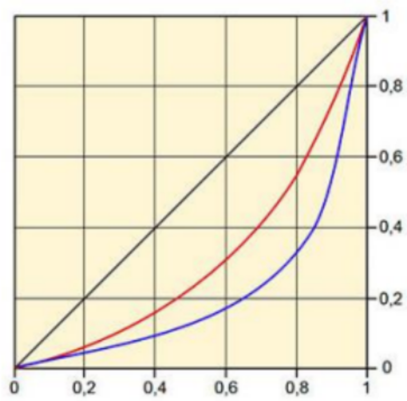


Figura 9. Exemple Corba de Lorenz. (Peña, 2019)

Veiem que hi ha representades la recta $y = x$ en negre, i dues corbes v (vermella) i b (blava), que podrien ser les Corbes de Lorenz que representin les desigualtats salarials anuals en dos països diferents. Com que la corba blava està més allunyada respecte de la diagonal, voldrà dir que els salaris de la població d'aquest país estan més descompensats que dels del país representat per la corba vermella.

La recta diagonal representa la distribució perfectament igualitària dels salaris, i va del punt (0,0), que representaria el 0% de percentatge acumulatiu, al punt (1,1), que representa el 100% del gràfic.

Com s'ha dit abans, l'índex de Gini es calcularia de la següent manera:

$$\begin{aligned} \textit{Índex de Gini} &= \frac{A}{A+B} \\ \textit{on } A &= \textit{àrea entre la corba i la diagonal}, \\ B &= \textit{àrea per sota de la corba}. \end{aligned}$$

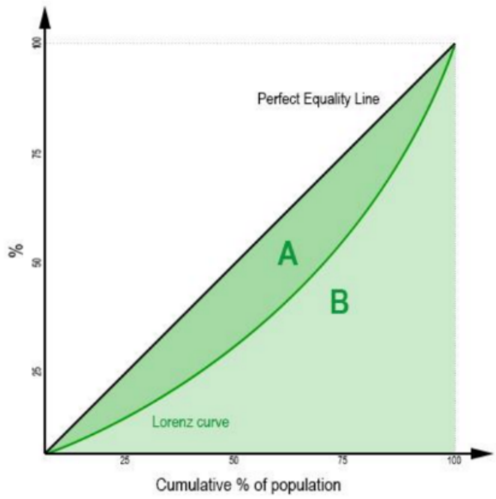


Figura 10. Àrees pel càlcul de l'índex de Gini. (Zeder, 2018)

Aquest coeficient té prou aplicabilitat en el risc de crèdit, ja que serà de molt interès quantificar les desigualtats de les probabilitats estimades pels models segons si l'individu ha complert els seus pagaments o no. De fet, contrari al que intuïtivament es pugui pensar, es buscaran models que facin augmentar l'índex de Gini, ja que com més gran sigui, més desigualtats hi haurà en les estimacions de les probabilitats de les observacions, i per tant, el model identificarà millor les dues classes d'interès: si el client farà *Default* o no.

4.4 Factors de risc

Són necessàries unes bases que regulin i tinguin en compte tots els possibles aspectes d'un model, quantifiquin tots els riscos i beneficis, i que serveixin al desenvolupador que està creant l'algoritme del model per a saber si existeix algun aspecte que es pugui millorar. Aquestes bases poden tenir en compte un gran nombre de criteris, i en el sector del risc de crèdit, majoritàriament es fan servir els *sistemes IRB (Internal Rating Based)*, que identifiquen i classifiquen tots els factors que poden suposar un cost per als desenvolupadors.

Els sistemes de qualificació IRB els proposa el *BdE (Banc d'Espanya)* i són models estadístics que avaluen internament el perfil de risc, i proposa unes restriccions que han de complir. Les validacions són principalment quantitatives, tot i que existeixen altres factors que recullen els sistemes IRB, com ara la privacitat de les dades i la seva qualitat i veracitat. S'utilitza molt per a l'estimació de les probabilitats d'impagaments, i són les institucions creditícies les responsables quan se sotmeten a l'avaluació d'aquests sistemes. Un cop hagi estat aprovat el disseny d'un model i compleixi amb tots els requisits supervisats, aquest és acceptat i pot ser utilitzat, per exemple, per a la predicció del capital a reservar necessari per a que una institució cobreixi correctament els seus riscos.

Sabent això, és important que els models de Machine Learning siguin compatibles amb els sistemes de validació IRB. Per a estudiar-ne aquesta compatibilitat i quantificar-ne els costos potencials, es defineixen quatre grans grups de factors de risc (o factors de costos) que permetran analitzar els inconvenients de cada model:

1. L'estadística: Aquests factors poden afectar en la creació del model de Machine Learning de forma directa, per exemple amb la presència d'hiperparàmetres, que són paràmetres ajustables a l'hora d'entrenar els models, o processar les dades, contrastar el comportament de les observacions...

2. La tecnologia: Que hi hagi una capacitat tecnològica suficient és absolutament necessari en aquest tipus de models, ja que no només s'han d'arribar a desenvolupar sinó que també s'han de poder executar amb un temps raonable i mantenir-se en producció per a la seva operació.

3. La interpretabilitat: És molt important tenir la capacitat d'interpretar el model i els seus resultats, per exemple en el cas del risc creditici, s'han de poder donar explicacions raonables als clients als quals se'ls hi denega o accepta un crèdit. De fet, la Comissió Europea estableix que els resultats dels models de Machine Learning han de ser interpretables per a totes les persones participatives en el procés, tant l'equip tècnic de desenvolupadors del model com el propi client, ja que aquest tipus de decisions poden afectar significativament l'impacte econòmic d'una persona o negoci particular. Així doncs, el banc ha de ser capaç d'explicar quins han estat els elements de risc, variables, o qualsevol altre aspecte que hagi determinat la decisió.

4. Els biaixos: S'han de complir els principis d'igualtat; totes les dades i preses de decisions han de ser lliures d'injustícies, estigmatitzacions i discriminacions. S'ha d'evitar que les mostres inicials estiguin ja esbiaixades, i en cap cas les decisions poden estar basades en característiques personals. També es fa referencia al biaix algorítmic, que seria que el model utilitzés únicament un petit conjunt de variables per a realitzar les prediccions.

El propòsit amb el que es construeix un model és important per a quantificar-ne els seus costos i beneficis. Per exemple, si s'està fent una predicció del capital a reservar pel banc, la precisió del model és fonamental, mentre que si l'objectiu és predir les probabilitats d'impagament (PD), el més important és la capacitat de classificació.

En general i en mode de resum, els models de Machine Learning són populars per tenir una molt bona capacitat predictiva de qualsevol problema. A més, a mesura que s'augmenta la complexitat de l'algoritme, per exemple amb els *Random Forest* o les *Xarxes Neuronals*, milloren les capacitats de predicció respecte dels models estadístics simples tradicionals. Per a seguir avançant en aquest terreny amb aparentment infinitat de millores, seria necessari realitzar encara més estudis utilitzant algoritmes d'*aprenentatge per reforç* o *xarxes neuronals convolucionals* per seguir millorant les prediccions.

Conclusions

Abans d'endinsar-me per primer cop en aquest món, el del risc creditici, no era conscient de que fos una especialització que donés abast a tantes coses. Donava per suposat que les finances i, en concret, els riscos financers eren un sector que m'interessava molt i del que s'hi podien aprendre moltes coses, però no va ser fins a entrar en l'estudi del risc de crèdit, que me'n vaig adonar que cada una de les seves branques era molt més ample del que pensava.

Amb aquest treball, he estudiat amb profunditat aquest camp. He establert primer unes bases teòriques econòmiques i financeres; després, he presentat uns conceptes matemàtics de càlcul estocàstic; més endavant, he estudiat els models de risc de crèdit tradicionals, fent servir termes i definicions dels capítols anteriors; i finalment, m'he endinsat breument en un capítol més modern, el de models de Machine Learning.

Així doncs, el principal objectiu de l'estudi era personal, i s'ha complert: Aprendre els conceptes principals del risc creditici, tant teòrics com pràctics, i establir un coneixement base per a entendre les parts més tècniques del tema. A més, aquestes eren un tipus de matemàtiques en les que considero que no m'hi havia endinsat prou durant el grau, ja que disten de la geometria, topologia, o àlgebra que he vist tan sovint, i s'endinsen una mica més en una part molt aplicable al món financer, la modelització.

Aquest apropament em permet relacionar dos camps tangencials de les matemàtiques que m'interessen molt: l'economia i l'algorítmica. Estic molt satisfet d'haver tingut l'oportunitat d'estudiar-ho i també crec que em dóna una experiència molt útil per al món laboral.

A més, he pogut reforçar el coneixement de càlcul estocàstic i modelització, que havia vist en poques ocasions durant el grau, i que m'han resultat molt útils d'aprendre i desenvolupar.

Com a oportunitat de millora, m'hauria agradat continuar el treball veient una vessant més pràctica de cada model que explico de manera teòrica, però això excedia el propòsit d'aquest treball, ja que considero que llavors no hagués pogut donar abast a tots els tipus de models que volia. Així doncs, en un context diferent i com a continuació natural del treball, podria establir uns exemples de problemes que pugui resoldre de manera pràctica amb programació. Això no només enriquiria el treball, sinó que laboralment em seria molt útil a nivell personal, ja que em permetria tenir una experiència fonamental i molt aplicable a problemes d'empreses reals.

Bibliografia

[1] Alonso, A; Carbó, J. (2020): *Machine Learning in Credit Risk: Measuring the dilemma between prediction and supervisory cost.* Banc d'espanya, 2020.

[2] Comité de Supervisión Bancaria de Basilea (2017): *Basilea III: Finalización de las reformas poscrisis*

[3] Baudoin, F(2012): Lecture 8: The Doob's stopping theorem. Disponible en: https://fabricebaudoin.wordpress.com/2012/03/26/lecture-8-the-doobs-stopping-theorem/ [Consulta: Novembre de 2022]

[4] Baudoin, F(2012): Lecture 11: Doob's martingale maximal inequalities. Disponible en: https://fabricebaudoin.wordpress.com/2012/03/26/lecture-8-the-doobs-stopping-theorem/ [Consulta: Novembre de 2022]

[5] BBVA(2015): IRFS9 Impact on the Basel Capital Framework. Disponible en: https://www.aeca.es/old/faif/presentaciones6/ifrs9_jv.pdf [Consulta: Novembre de 2022]

[6] Bielecki, T; Rutkowski, M. (2002): *Credit Risk: Modelling, Valuation and Hedging.* Springer Finance.

[7] Brigo, D; Mercurio, F. (2006): *Interest Rate Models - Theory and Practice.* Springer Finance.

[8] Costa, T; Boj, E; Fortiana, J. (2012): *Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias. Aplicación al problema de Credit Scoring*

[9] Gikham, I; Skorhod, A. V. (1969): *Introduction to the theory of random processes.* Traduït del Rus a l'anglès per Scripta Technica, Inc. W. B. Saunders Co.

[10] Jacod, J; Protter P. (2003): *Probability Essentials.* Springer.

[11] Karatzas, I; Shreve, S. (1988): *Brownian Motion and Stochastic Calculus.* Springer.

[12] Lamberton, D; Lapeyre, B. (2008): *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall / CRC.

[13] Martos, A. (2021): *Anàlisi de les diferents eines de modelització en el camp del risc de crèdit* UPCommons. Global access to UPC knowledge.

[14] Maximenko, E: Teorema de la convergencia monótona. Disponible en: https://esfm.egormaximenko.com/analysis/monotone_convergence_theorem_es.pdf [Consulta: Octubre de 2022]

[15] Narkhede, S. (2018): *Understanding Confusion Matrix. Medium.* Journal of Data Analysis and Information Processing.

[16] Nelsen, R. B. (2006): *An Introduction to Copulas.* Springer.

[17] Peña, P. (2019): Que es la Curva de Lorenz. Disponible en: https://es.scribd.com/document/367796672/Que-Es-La-Curva-de-Lorenz [Consulta: Desembre de 2022]

[18] Rudloff, B. (2014): *Quantitative Risk Management Lecture.* WU Vienna.

[19] Wikipedia: Corba ROC. Disponible en: https://es.wikipedia.org/wiki/Curva_ROC [Consulta: Desembre de 2022]

[20] Wikipedia: Coeficient de Gini. Disponible en: https://es.wikipedia.org/wiki/Coeficiente_de_Gini [Consulta: Desembre de 2022]

[21] Wikipedia: Linear Regression. Disponible en: https://es.wikipedia.org/wiki/Regresion_lineal_simple [Consulta: Desembre de 2022]

[22] Wikipedia: Logistic Regression. Disponible en: https://es.wikipedia.org/wiki/Regresion_logistica [Consulta: Desembre de 2022]

[23] Zeder (2018): Quantitative Measures of Wealth Inequality in Ancient Central Mexican Communities. Disponible en: https://www.researchgate.net/publication/282914616_Quantitative_Measures_of_Wealth_Inequality_in_Ancient_Central_Mexican_Communities [Consulta: Desembre de 2022]

[24] Zurita, J; Rodríguez, F; Martínez, J. (2009): *La crisis financiera y económica del 2008. Origen y consecuencias en los Estados Unidos y México.* El Cotidiano, Septiembre-Octubre, 17-27.