

Distribuciones de los estadísticos y estimadores paramétricos de la Regresión Ancestral bajo apareamiento aleatorio y selección

*Tesis presentada para optar al título de Magister de la Universidad de Buenos Aires,
Área Biometría y Mejoramiento*

Esteban Jiménez Alfaro
Lic. Ing. Agr. Zootecnista - Universidad de Costa Rica - 2006

Lugar de Trabajo: Universidad Nacional de Costa Rica
Facultad de Ciencias de la Tierra y el Mar - Escuela de Ciencias Agrarias



Escuela para Graduados Ing. Agr. Alberto Soriano
Facultad de Agronomía - Universidad de Buenos Aires



COMITÉ CONSEJERO

Director de tesis

Rodolfo Juan Carlos Cantet

Ing. Agr. (Universidad de Buenos Aires, Argentina)

MSc. (Montana State University, E.E.U.U.)

MSc. (University of Illinois, E.E.U.U.)

Ph.D. (University of Illinois, E.E.U.U.)

Co-director

Natalia Soledad Forneris

Ing. Agr. (Universidad de Buenos Aires, Argentina)

Doctor (Universidad de Buenos Aires, Argentina)

JURADO DE TESIS

JURADO

Daniel Omar Maizón

Med. Vet. (Universidad de Buenos Aires, Argentina)

MSc. (Universidad de Buenos Aires, Argentina)

Ph.D. (Cornell University, E.E.U.U.)

JURADO

Eduardo Néstor Fernández

Ing. Zootecnista (Universidad Nacional de Lomas de Zamora)

Dr. (Universidad Politécnica de Valencia)

JURADO

María del Carmen Fabrizio

Prof. Mat. (Instituto Sagrado Corazón, Argentina)

Lic. Cs. Mat. (Universidad de Buenos Aires, Argentina)

MSc. Biometría (Universidad de Buenos Aires, Argentina)

Dra. Cs. Biol. (Universidad de Buenos Aires, Argentina)

Fecha de defensa de la tesis: 2 de agosto de 2019

A mi familia.

AGRADECIMIENTOS

Agradezco primero a Dios y a la Virgen María por mantenerme siempre en las palmas de sus manos.

Especialmente al comité consejero: Fito y Natalia, quienes sobre todo me enseñaron el valor del esfuerzo y la perseverancia en el trabajo.

A los jurados, quienes dedicaron su tiempo desinteresadamente y quienes, con sus comentarios constructivos ayudaron a mejorar esta tesis.

A la querida profesora María del Carmen por su especial apoyo y enseñanza durante todo mi proceso de aprendizaje.

A mis amigos de la Cátedra de Mejoramiento Genético Animal de la FAUBA: Sebas, Andrés, Caro, Matías, Belcy, Roberto, Martín y Dani por su apoyo incondicional y sabios consejos.

¡A mi gran amigo Gastón! por todo su apoyo y compañía.

A la Escuela para Graduados, muy especialmente a Susi y Mariano.

A mi familia por su cariño, comprensión y constante amor.

Finalmente, a todas aquellas personas a quienes tuve la bendición de conocer.

DECLARACIÓN

Declaro que el material incluido en esta tesis es, a mi mejor saber y entender, original producto de mi propio trabajo (salvo en la medida en que se identifique explícitamente las contribuciones de otros), y que este material no lo he presentado, en forma parcial o total, como una tesis en esta u otra institución.

Esteban José Jiménez Alfaro

ÍNDICE GENERAL

DEDICATORIA	iii
AGRADECIMIENTOS	iv
DECLARACIÓN	v
ÍNDICE GENERAL	vi
ÍNDICE DE CUADROS	viii
ÍNDICE DE FIGURAS	x
ABREVIATURAS	xi
RESUMEN	xii
ABSTRACT	xiii
CAPÍTULO 1. Introducción	1
CAPÍTULO 2. Teoría de la Regresión Ancestral.....	6
2.1. La Regresión Ancestral (AR)	7
2.2. Estimación de los parámetros β	12
2.3. Cálculo de la fracción de genoma compartido IBD	18
CAPÍTULO 3. Materiales y Métodos	20
3.1. Metodología	21
3.2. Base de datos simulados	22
3.3. Estimación de las densidades de las relaciones de IBD compartido abuelo nieta (RIBD-AN) y de los parámetros de Regresión Ancestral (AR)	23
3.4. Relaciones de parentesco	27
3.4.1. Relaciones esperadas	27
3.4.2. Relaciones realizadas	28
3.5. Estimación de la matriz de covarianzas Σ	28
3.5.1. Matriz Σ de orden 3	28
3.6. Consideración de los efectos de la selección	31
3.7. Coeficiente de endogamia (F_{AR})	32
3.8. Estimación de los parámetros β_S , β_D y F_{AR}	33
3.9. Análisis estadístico	33
CAPÍTULO 4. Resultados	34
4.1. Distribución de las Relaciones de IBD compartido (RIBD)	35

4.1.1. Parámetros de la distribución Beta	36
4.1.2. Estimadores de los momentos de la distribución Beta para RIBD-AN	36
4.1.3. Curtosis, asimetría y otros estadísticos de las distribuciones empíricas	37
4.1.4. Distribuciones empíricas por generación	39
4.1.5. Distribuciones de los estadísticos suficientes	41
4.2. Parámetros de la Regresión Ancestral (AR)	42
4.2.1. Coeficiente de regresión parcial β_s y β_D	42
4.2.2. Consanguinidad (F_{AR})	43
4.2.3. Correlaciones entre las RIBD-AN y los parámetros β_S , β_D y F_{AR}	44
CAPÍTULO 5. Discusión	47
5.1. Relaciones de IBD compartido abuelo-nieto	48
5.2. Parámetros del modelo de Regresión Ancestral (AR)	50
CAPÍTULO 6. Conclusiones	52
BIBLIOGRAFÍA	55

ÍNDICE DE CUADROS

Cuadro 1. Verosimilitudes de distribuciones ajustadas por Máxima Verosimilitud, para las RIBD-AN, estimadas bajo AR sobre RM y SEL	35
Cuadro 2. Parámetros de la distribución Beta con sus respectivos errores estándar (EE), para las RIBD-AN, estimadas bajo AR, sobre RM y SEL	36
Cuadro 3. Estimadores de las esperanzas $E(X)$, varianzas $Var(X)$, medias muestrales \bar{X}_n , varianzas muestrales S_x^2 , desvíos estándar DS, y coeficientes de variación CV, para las RIBD-AN estimadas bajo AR bajo apareamiento aleatorio RM o selección SEL	37
Cuadro 4. Mediana, media, desvío estándar, mínimo, máximo, asimetría y curtosis para las RIBD-AN, estimadas bajo AR sobre RM y SEL	38
Cuadro 5. Media, desvío estándar y coeficiente de variación por generación para las distribuciones de las RIBD-AN, estimadas bajo AR sobre RM y SEL	41
Cuadro 6. Mínimo, máximo, mediana, media, desvío estándar, asimetría y curtosis de las distribuciones de los estadísticos suficientes, estimadas bajo AR sobre RM y SEL	42
Cuadro 7. Media y desvío estándar de los coeficientes de regresión parcial: paterno (β_S) y materno (β_D), estimados bajo AR sobre RM y SEL	43
Cuadro 8. Parámetro de la distribución exponencial (λ) y su error estándar para los valores de consanguinidad estimadas por pedigrí (F_{PED}), empleando información genómica (F_{GENO}) y por Regresión Ancestral (F_{AR}), según los datos sean RM o SEL	44

Cuadro 9. Media y Varianza de los valores de F_{PED} , F_{GENO} y F_{AR} para datos RM y SEL, asumiendo distribuciones exponenciales para los parámetros estimados	44
Cuadro 10. Coeficientes de correlación entre RIBD-AN, coeficientes de regresión parcial paterno (β_{S}) y materno (β_{D}) y coeficiente de consanguinidad (F_{AR}) bajo AR sobre RM	45
Cuadro 11. Coeficientes de correlación entre RIBD-AN, coeficientes de regresión parcial paterno (β_{S}) y materno (β_{D}) y coeficiente de consanguinidad (F_{AR}) bajo el AR sobre SEL	46

ÍNDICE DE FIGURAS

- Figura 1.** Diagrama de dos individuos que comparten segmentos idénticos por descendencia IBD, proveniente de un ancestro en común. Los cuadrados representan machos y los círculos hembras. Cada barra dentro de una figura representa la gameta paterna y materna 18
- Figura 2.** Pedigrí incluyendo hasta los tatarabuelos del individuo X. Los distintos nodos representan a sus padres (2 meiosis), los abuelos paternos y maternos (4 meiosis), los bisabuelos paternos y maternos (8 meiosis) y finalmente, los tatarabuelos paternos y maternos (16 meiosis) 31
- Figura 3.** Gráfico de la función de densidad según RIBD-AN: SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna), estimadas bajo el modelo de Regresión Ancestral, sobre: **a.** Datos bajo selección (SEL) y **b.** Datos bajo apareamiento aleatorio (RM). Las líneas verticales para SEL fueron mediana = 0.3175, media = 0.3233; y para RM, mediana = 0.2676 y media = 0.2725 39
- Figura 4.** Funciones de densidad por RIBD-AN, donde SS es el abuelo paterno, DS es la abuela paterna, SD es el abuelo materno y DD es la abuela materna, estimadas por el modelo de Regresión Ancestral, sobre datos SEL y RM, para cada una de las generaciones comprendidas en cada réplica 40

ABREVIATURAS

AR	Regresión Ancestral (del inglés <i>Ancestral Regression</i>)
BLUP	Predicción lineal insesgada de mínima varianza (del inglés <i>Best Linear Unbiased Prediction</i>)
CV	Coefficiente de variación
DD	Abuela materna (del inglés <i>maternal grand dam</i>)
DS	Desvío estándar
DS	Abuela paterna (del inglés <i>paternal grand dam</i>)
HMM	Modelos Ocultos de Markov (del inglés <i>Hidden Markov Models</i>)
IBD	Identidad por Descendencia (del inglés <i>Identity by Descent</i>)
IBS	Identidad por estado (del inglés <i>Identity by State</i>)
LD	Desequilibrio de Ligamiento (del inglés <i>Linkage Disequilibrium</i>)
MGP	Abuelo Materno (del inglés <i>Maternal Grand-Parents</i>)
MME	Ecuaciones de Modelo Mixto (del inglés <i>Mixed Model Equations</i>)
PGP	Abuelo Paterno (del inglés <i>Paternal Grand-Parents</i>)
RIBD	Relaciones IBD compartido
RIBD-AN	Relaciones IBD compartido abuelo-nieto
RM	Apareamiento Aleatorio (del inglés <i>Random Mating</i>)
ROH	Corridas de homocigosidad (del inglés <i>Runs of Homozygosity</i>)
SD	Abuelo materno (del inglés <i>maternal grand sire</i>)
SS	Abuelo paterno (del inglés <i>sire of sire</i>)
SEL	Conjunto de datos bajo selección
SNP	Polimorfismos genéticos de un solo nucleótido (del inglés <i>Single Nucleotide Polymorphysm</i>)
VC	Valor de Cría (del inglés <i>Breeding Value</i>)

RESUMEN

Distribuciones de los estadísticos y estimadores paramétricos de la Regresión Ancestral bajo apareamiento aleatorio y selección

La Regresión Ancestral (AR) es un modelo genético cuantitativo paramétrico y causal. A diferencia de otros modelos de predicción basados en Selección Genómica, AR permite considerar el efecto de la recombinación y la herencia en segmentos, así como realizar la inversión de la matriz de (co)varianzas del modelo (Σ^{-1}) sin necesidad de calcular primero Σ , porque además AR es un proceso estocástico Markoviano. El objetivo de esta tesis fue estimar mediante simulación estocástica las distribuciones de probabilidad de: a) los parámetros del modelo AR (β_S , β_D), b) las relaciones realizadas de identidad por descendencia (IBD) entre pares abuelo(a)-nieto(a) (RIBD-AN), c) los estadísticos suficientes para la estimación de β_S y β_D , y d) el coeficiente de consanguinidad bajo AR (F_{AR}). Se simularon dos poblaciones animales, una con apareamientos al azar (45095 pares) y otra bajo selección (46561 pares). El algoritmo de estimación incluyó un sistema recursivo de ecuaciones simultáneas de los parámetros β_S y β_D , los cuales provienen de las recombinaciones del genoma de los abuelos ocurridas en los padres durante la meiosis y transmitidas a los nietos. Se estimaron los “núcleos” (o kernel) de las distribuciones de RIBD-AN, los estadísticos suficientes, β_S , β_D y F_{AR} con un método no paramétrico dentro del programa PROC KDE de SAS. Entre distintas distribuciones evaluadas, la Beta mostró el mejor ajuste para las RIBD-AN, tomando como criterio el valor máximo del logaritmo de la función de verosimilitud; en tanto que los estimadores de β_S , β_D y los estadísticos suficientes siguieron distribuciones normales univariadas. La selección indujo a las distribuciones de las RIBD-AN a ser asimétricas a la derecha y a mostrar mayor magnitud en las medias y en la dispersión. Además, se observó que F_{AR} siguió una distribución Exponencial, con valores esperados y varianzas de mayor magnitud con los datos selectos. En general, la selección aumentó la magnitud y variabilidad del material genómico compartido IBD, tanto en las relaciones de parentesco como en la consanguinidad.

Palabras clave: Valor de cría, inferencia causal, herencia en segmentos, proceso Markoviano, datos genómicos, relaciones de parentesco IBD compartido abuelo nieto.

ABSTRACT

Distributions of the statistics and parametric estimators of the Ancestral Regression under random mating and selection

The Ancestral Regression (AR) is a parametric and causal quantitative genetic model that allows considering the effect of recombination and inheritance in segments, also allowing an inversion of the AR process matrix (Σ^{-1}) without first calculating Σ , since AR is also a Markovian stochastic process. By means of stochastic simulation we estimated the probability distributions of: a) the parameters of the Ancestral Regression (β_S, β_D), b) the realized relationships of identity by descendent (IBD) for grandparent - grandprogeny pairs (RIBD-AN), c) sufficient statistics for the estimation of β_S and β_D , and d) the inbreeding coefficient under AR (F_{AR}). Two animal populations were simulated under random mating (45095 pairs) or selection (46561 pairs). The estimation algorithm included a recursive system of simultaneous equations for the parameters β_S and β_D , which originate in recombinations from the genome of the grandparents while occurring in the parents during the meiosis, to finally form the grand-progeny genome. The kernel of the RIBD-AN distributions, the sufficient statistics, β_S, β_D and F_{AR} were estimated using a non-parametric method from SAS PROC KDE. Among the different distributions that were evaluated, the Beta showed the best fit for RIBD-AN. The criterion to decide the impact of the fit was the maximum value of the logarithm of the likelihood function. The estimators of β_S, β_D and the sufficient statistics followed normal univariate distributions. Selection induced that the distributions of the RIBD-AN be asymmetric to the right and to show greater magnitude in the means and in the dispersion. In addition, it was observed that F_{AR} followed an Exponential distribution, with expected values and variances of greater magnitude for the selected data. Regarding inbreeding, F_{AR} showed intermediate estimated values between the inbreeding estimates obtained by pedigree and those generated using IBD genomic information. In general, selection increased the magnitude and variability of the proportion of genome shared IBD, either for the realized relationships or inbreeding.

Key words: breeding value, causal inference, segmental inheritance, Gaussian Markov density, genomic data, relationships of identity by descendent grandfather grandson.

Capítulo 1: Introducción

Capítulo 1

Introducción

La genética cuantitativa es una ciencia aplicada que tiene por objeto teorizar y cuantificar el efecto de la herencia de caracteres complejos, gobernados por la acción de muchos genes. La base teórica está en el trabajo de Fisher (1918) y fue extendida por la escuela de Edimburgo (Alan Robertson, Douglas Falconer, William Hill) popularizada en el texto de Falconer (1960). En este libro, se postula al valor de cría (VC) como el resultado de la suma de efectos génicos *independientes* de muchos loci, cada uno con efecto pequeño. Si bien postula la acción de genes individuales, el modelo *aditivo* no describe cómo identificar los genes causales con los fenotipos y apela a distribuciones *muestrales* (sampling distributions). A partir de 1930, desde el otro lado del Atlántico, se desarrolló la teoría de predicción del VC, con los trabajos de J. L. Lush (1935, 1945, 1964) y, más formalmente, Hazel (1943), ambos trabajando en Iowa State University. Este cuerpo teórico emplea una distribución normal implícita para el VC del individuo (identificación), sin necesidad de identificar los efectos de los genes. La teoría de predicción del VC fue ampliada y generalizada, para considerar en el análisis que los distintos valores esperados de los fenotipos debían ser estimados, por un discípulo de J. L. Lush: C. R. Henderson, quien desarrolló la teoría de modelos lineales mixtos para la predicción lineal insesgada de mínima varianza (BLUP) (Henderson, 1963, 1973, 1975). De particular interés al tema de esta tesis es la predicción del VC mediante el “modelo animal” (Henderson y Quaas, 1976), donde aparece por primera vez la noción de “regresar” el VC individual al promedio de los VC parentales, sumándole al modelo un error de transmisión hereditaria: el residuo Mendeliano donde la mitad (o algo menos con consanguinidad) de la varianza de los VC queda sin explicar. Sin que C. R. Henderson supiese esta “regresión parental” introduce una distribución *causal* (Pearl, 2000) y Markoviana, que simplifica el cálculo predictivo del VC. El modelo animal permitió resolver la inversa de la matriz de relaciones aditivas (A^{-1}) sin nunca tener que formar primero A , matriz necesaria para calcular las ecuaciones BLUP de C. R. Henderson. La técnica dominó enteramente la predicción del VC animal y vegetal y aún hoy es ampliamente utilizada.

El “modelo animal” Gaussiano como paradigma predictivo de la genética cuantitativa, iba a ser desafiado con el advenimiento de los paneles densos de marcadores moleculares a principios del nuevo milenio. Resurge entonces el modelo de predicción del VC como la suma de las acciones de efectos de polimorfismos de una sola base (SNP) conocida como “selección genómica”, luego del trabajo de Meuwissen et al. (2001). Esencialmente, esta propuesta consiste en regresar funciones lineales de fenotipos que no dependen de los efectos fijos en los SNP. Con un éxito comercial sin precedentes en la historia de la mejora genética animal, la selección genómica se vuelve una herramienta predictiva potente porque anticipa el momento de selección, especialmente en lechería donde los toros son ahora selectos a los pocos meses de edad, logrando un aumento del diferencial de selección y un fuerte acortamiento del intervalo generacional (Schaeffer, 2006). Desde un punto de vista teórico, la “regresión entera del genoma” (WGR) de T. Meuwissen, B. y M. Goddard sufrirá profundas críticas en cuanto hace al modelo predictivo del VC (Gianola et al., 2009; De los Campos et al., 2015) dado que los “marcadores no son genes”. Más que un modelo de regresión, WGR es un “modelo de error de medición” o “error en las variables” predictivas (Fuller, 1987; Buonaccorsi, 2010) que tiende a sobrestimar enormemente la exactitud de las predicciones del VC obtenidas y explica menos varianza de los VC que el “modelo animal”, fenómeno conocido como la “heredabilidad perdida” (“missing heritability”). Posteriormente, Legarra et al. (2009) transforman WGR agregando la información del pedigrí en la matriz A para poder emplear BLUP-“modelo animal”, transformado la ecuación predictiva del VC de causal a correlacional o multicolineal. Esto ocurre porque el VC individual es regresado en los VC de todos sus parientes de cualquier generación. Dicho método, conocido como BLUP en “un solo paso” (ssGBLUP del inglés single step BLUP), es muy utilizado en la actualidad, pero teóricamente es más un intento de generar una matriz de covarianzas “actualizada” por los marcadores, que un modelo del proceso de la herencia de los VC.

Así como en la definición del VC por la escuela de Edimburgo o la noción predictiva Gaussiana de Lush-Henderson, ssGBLUP descansa en la distribución normal multivariada de acciones de genes *independientes*: el “modelo infinitesimal” (Lynch y Walsh, 1998). Sin embargo, el genoma se transmite en segmentos (ligamiento físico de los genes en un mismo cromosoma, Schnell, 1961) y, además, existe *ausencia de independencia* estadística entre los efectos de genes o de marcadores: desequilibrio gamético o de ligamiento (LD) de genes en cromosomas distintos o en el mismo

cromosoma (Lynch y Walsh, 1998). Desde un punto de vista predictivo, es posible modelar la herencia del VC considerando esta ausencia de independencia de las acciones génicas o herencia en segmentos, sobre la base de asociar la distribución causal de la herencia a la noción de identidad por descendencia (IBD) (Malécot, 1969). Más aún, J. B. S. Haldane (Cockerham y Weir, 1969) demostró que la probabilidad conjunta en dos loci muestreados en cualquier cromosoma de dos individuos emparentados sean IBD es *siempre mayor* que el producto de las probabilidades marginales IBD de cada uno de dichos loci. En otras palabras, una matriz de relaciones entre individuos (dígase Σ) que funcione como matriz de covarianzas del VC, tendrá elementos mayores que A y, cuando la diferencia $\Sigma - A$ sea positiva definida, el predictor BLUP(VC) obtenido recuperará mayor varianza de los VC y heredabilidad que con el modelo animal (Cantet y Vitezica, 2014). Este argumento fue utilizado por Cantet et al. (2017) para generar el modelo de “Regresión Ancestral” (AR). Para ello emplearon una distribución *causal* de la herencia del VC, donde el VC individual se origina o es causado por el VC ancestral. Como en Lange (1978), Bulmer (1985) y Abney et al. (2000) AR es Gaussiano. También, como en Henderson y Quaas (1976) AR es un modelo Markoviano porque condicional al VC de padres y abuelos, el residuo Mendeliano de un individuo en AR es independiente de cualquier otro residuo Mendeliano. Para tal fin, AR emplea dos parámetros que son coeficientes de regresión parcial del VC individual en la *diferencia de los VC de los abuelos paternos* (β_s) y *maternos* (β_D). Tanto β_s como β_D cuantifican las *recombinaciones* ocurridas durante las meiosis originarias que crean al individuo, e indican el exceso o defecto del genoma de un abuelo respecto del genoma de la abuela, por encima o debajo de 25% del genoma del animal. Dichos parámetros son identificables solo mediante paneles densos de marcadores SNP o secuencias de ADN. Por lo tanto, el ajuste del modelo AR a la evaluación genética animal requiere estimar β_s y β_D previo al cálculo de Σ^{-1} para obtener el BLUP(VC), empleando información del pedigrí y de paneles densos de ADN. Para estimar los parámetros con métodos de verosimilitud o Bayesianos, es necesario entonces conocer la distribución de los valores estimados de β_s y β_D , así como también la de las relaciones genómicas abuelo(a) – nieto(a) (RIBD-AN) que son *estadísticos suficientes* del modelo AR (Cantet et al., 2017). Dada la dificultad teórica de modelar las distribuciones de probabilidad exactas para procesos nunca antes observados ni medidos, se adoptará un enfoque empírico mediante simulación de Monte Carlo para aproximar estas distribuciones, particularmente en presencia de selección, dado el

impacto que tiene este proceso sobre las predicciones del VC (Henderson, 1975; Im et al., 1989; Gianola et al., 2009). Consecuentemente, los objetivos generales de esta tesis son: 1. evaluar distintas distribuciones continuas para las relaciones de identidad por descendencia abuelo-nieto (RIBD-AN) bajo AR; 2. obtener las distribuciones de los coeficientes de regresión parcial paterno (β_s) y materno (β_D), así también como la distribución de los estadísticos suficientes; 3. Evaluar el efecto de la *selección* sobre las RIBD-AN, y finalmente, 4. Obtener la distribución del coeficiente de consanguinidad bajo AR (F_{AR}), de modo de poder calcular los elementos diagonales de Σ .

El documento está organizado en seis capítulos siendo el primero está introducción. El Capítulo 2 contiene el marco teórico, donde se describe la Regresión Ancestral, así como también el método de estimación de β_s y β_D . Se informa allí sobre a) la *estimación de la fracción del genoma compartido* IBD, b) la *estimación de las distribuciones* de las relaciones IBD ancestro-individuo, de los coeficientes de regresión parcial y del coeficiente de consanguinidad. El Capítulo 3 describe la metodología utilizada para el proceso de simulación; se detallan las estimaciones de las funciones de densidad de las relaciones de IBD compartido abuelo-nieto (RIBD-AN) y de los parámetros del modelo AR, bajo apareamiento aleatorio (RM) empleado como “control”, y bajo selección (SEL). Además, en este capítulo se describen el coeficiente de consanguinidad bajo AR, el algoritmo utilizado para la estimación de los parámetros y el análisis estadístico utilizado para ajustar las distribuciones empíricas. Los resultados obtenidos se presentan en el Capítulo 4. El Capítulo 5 contiene la discusión de los resultados y, finalmente el Capítulo 6 puntualiza las conclusiones generales de la tesis.

Capítulo 2: Teoría de la Regresión Ancestral

Capítulo 2

Teoría de la Regresión Ancestral

2.1. La Regresión Ancestral (AR)

La Regresión Ancestral (AR) es un modelo genético cuantitativo que generaliza la regresión del valor de cría (VC) en el promedio de los VC paternos (Henderson y Quaas, 1976; Bulmer, 1985) y que fuera propuesto por Cantet et al. (2017). Expande el concepto de “identidad por descendencia” (IBD) (Malécot, 1969), comúnmente utilizado para referir a genes muestreados aleatoriamente en el genoma de dos individuos que poseen un antecesor común, hacia segmentos del genoma. Desde un punto de vista estadístico deja de lado la noción de *independencia* en la transmisión de los efectos génicos (variables aleatorias) de padres a hijos para pasar a una versión más realista en la cual el genoma se transmite en segmentos por causa de la recombinación ocurrida durante la meiosis. Esto induce a que los efectos génicos no sean independientes de aquellos efectos causales originados en genes vecinos en el mismo segmento. Esta noción conduce a definir la relación de parentesco entre dos individuos como la proporción de genoma compartido IBD entre ambos animales (Guo, 1995). En tal sentido, AR modela el VC considerando las relaciones de parentesco “realizadas” debido a la recombinación, al ligamiento y al desequilibrio gamético (o de ligamiento, LD).

El modelo AR transforma un sistema recursivo de ecuaciones simultáneas en un modelo de regresión, que incluye como nuevos parámetros, los coeficientes de regresión parcial provenientes de las recombinaciones del material genético de los abuelos, ocurridas en los padres durante la meiosis y transmitidas al individuo. Dichos coeficientes, complementan la información aportada por el pedigrí con la brindada por un conjunto denso de marcadores genómicos, de manera tal, que la función lineal resultante del valor de cría de los abuelos, no se encuentra correlacionada con el promedio del VC de los padres en ausencia de consanguinidad. Más formalmente, AR puede ser descrito por el siguiente proceso Gaussiano (dado que la distribución de los VC es normal, Lange (1978)):

$$a_x = 0.5 a_s + 0.5 a_d + \beta_s (a_{ss} - a_{ds}) + \beta_d (a_{sd} - a_{dd}) + \phi_x \quad [2.1]$$

Los valores a_S y a_D corresponden al VC del padre (S) y de la madre (D) del individuo X; a_{SS} y a_{DS} son los VC del abuelo paterno (SS) y de la abuela paterna (SD) del individuo X y a_{SD} y a_{DD} refieren a los VC del abuelo materno (SD) y de la abuela materna (DD) de X; ϕ_X es el residuo Mendeliano, mientras que β_S y β_D son los parámetros de AR. Dichos parámetros reflejan la recombinación de los genomas de ambos pares de abuelos, que se presentan en las gametas parentales que constituyen al individuo, provocando un aumento (o una reducción) respecto de 0.25, o fracción de genoma esperado en un nieto por cualquier abuelo o abuela en ausencia de consanguinidad. Arbitrariamente, Cantet et al. (2017) fijaron que valores positivos de β_S y β_D corresponden a exceso de material genómico de los abuelos, mientras que signos negativos para los parámetros indican exceso de material genómico de las abuelas.

Los VC generados por el modelo de Regresión Ancestral (AR) mantienen una distribución causal Gaussiana, con una estructura de covarianzas Σ de características Markovianas. Esto quiere decir que, condicional a los VC de padres y abuelos el residuo Mendeliano de cualquier individuo es independiente de cualquier otro VC o residuo Mendeliano.

La matriz de covarianzas de los VC en AR es igual a $\Sigma\sigma_A^2$ (Cantet et al., 2017) siendo el escalar σ_A^2 la varianza genética aditiva. El elemento Σ_{XY} de la matriz estructural de covarianzas, es la relación de parentesco realizada u observada entre los individuos X e Y, y es igual a la fracción del genoma IBD compartido (denotado por “ \equiv ”) entre X e Y (Guo, 1995):

$$\Sigma_{XY} = P_c(g_x = g_y | g_c \downarrow g_x \uparrow g_y) = P(g_x \equiv g_y) \quad [2.2]$$

Los valores g_x y g_y son variables aleatorias que representan uno de los dos genomas parentales del individuo X el Y, mientras que g_c es el genoma de todos los ancestros comunes para ambos individuos. Las flechas indican el flujo del material genómico desde el ancestro común “C” hasta los individuos “X” e “Y”. Cantet, R. J. C. 2019, comunicación personal, observó que la expresión [2.2] corresponde a una variable aleatoria continua (Donnelly, 1983), que no es equivalente a la clásica versión discreta de la probabilidad de IBD entre dos o más genes (Bickeböllner y Thompson, 1996).

La herencia de los valores de cría entre generaciones y las covarianzas entre parientes se explican mediante dos expresiones (Cantet, R. J. C. 2019, comunicación personal). La

primera es para el caso en que ni X ni Y son descendientes del otro y proviene de la descomposición de Bartlett de una matriz (Munilla y Cantet, 2012):

$$\begin{bmatrix} \Sigma_A & \Sigma_A \mathbf{B}_Y \\ \mathbf{B}_X' \Sigma_A & \mathbf{B}_X' \Sigma_A \mathbf{B}_Y \end{bmatrix} \sigma_A^2 \quad [2.3]$$

El bloque Σ_A es la matriz cuadrada conteniendo la estructura de covarianzas entre los cuatro abuelos y los dos padres de X e Y, e incluye además los otros ancestros comunes a ambos individuos. El vector $\mathbf{B}_X' = [0 \dots \beta_s \quad -\beta_s \quad \dots 0 \dots \beta_D \quad -\beta_D \quad 0.5 \quad 0.5]$ contiene los parámetros de interés o coeficientes de regresión parcial (β) correspondientes al abuelo paterno (β_s), la abuela paterna ($-\beta_s$), el abuelo materno (β_D), la abuela materna ($-\beta_D$), el padre (S) (0.5) y la madre (D) (0.5), estando los valores cero en las columnas que hacen que la multiplicación $\Sigma_A \mathbf{B}_X = \Sigma_{AX}$ sea posible. El vector \mathbf{B}_Y tiene características similares a \mathbf{B}_X pero asociado a los parámetros beta, valores 0.5 y 0 propios de Y. Los vectores Σ_{AX} y Σ_{AY} contienen las covarianzas entre los ancestros en A con X e Y, respectivamente. Finalmente, $\mathbf{B}_X' \Sigma_A \mathbf{B}_Y = \Sigma_{XY}$ es igual a la relación entre los genomas compartidos IBD de X con Y. En caso de que X = Y, i.e. sea un único individuo dígame X, el elemento Σ_{XX} se halla en la diagonal de la matriz estructural, siendo igual a uno (1) más el coeficiente de consanguinidad bajo el modelo AR. Cantet (2019, inédito) obtuvo la siguiente expresión para la consanguinidad bajo AR:

$$F_{X(AR)} = \frac{1}{2} \left[\Sigma_{S,D} + \beta_s (\Sigma_{SS,D} - \Sigma_{DS,D}) + \beta_D (\Sigma_{S,SD} - \Sigma_{S,DD}) + \beta_s \beta_D (\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}) \right] \quad [2.4]$$

La varianza del residuo mendeliano de X en el modelo AR (Cantet et al., 2017) es:

$$\text{Var}(\phi_X) = \left[(1 + F_X) - \mathbf{B}_X' \Sigma_A \mathbf{B}_X \right] \sigma_A^2 = D_X \sigma_A^2 \quad [2.5]$$

donde el escalar D_X es el elemento diagonal de la matriz de covarianzas del residuo Mendeliano del VC de X.

La expresión más importante para obtener la covarianza entre VC es aquella entre un ancestro (A) y un descendiente (X; Cantet, R. J. C. 2019, comunicación personal), que permite obtener las covarianzas entre X y sus padres, abuelos y otros ancestros lejanos. Esta covarianza es igual a:

$$\text{cov}(a_A, a_X) = \Sigma_{A,X} = 0.5(\Sigma_{A,S} + \Sigma_{A,D}) + \beta_s (\Sigma_{A,SS} - \Sigma_{A,DS}) + \beta_D (\Sigma_{A,SD} - \Sigma_{A,DD}) \quad [2.6]$$

La expresión [2.6] permite obtener las siguientes covarianzas entre X y sus padres y abuelos:

Covarianza Padre – hijo: [2.6.A]

$$\text{cov}(a_s, a_x) = 0.5(1 + F_s + \Sigma_{s,D}) + \beta_s(\Sigma_{ss,s} - \Sigma_{ds,s}) + \beta_D(\Sigma_{sd,s} - \Sigma_{dd,s})$$

Covarianza Madre – hijo:

$$\text{cov}(a_D, a_x) = 0.5(1 + F_D + \Sigma_{s,D}) + \beta_s(\Sigma_{ss,D} - \Sigma_{ds,D}) + \beta_D(F_{SD} - F_{DD})$$

Covarianza Abuelo paterno – nieto:

$$\text{cov}(a_{ss}, a_x) = 0.5\Sigma_{ss,s} + 0.5\Sigma_{ss,D} + \beta_s(1 + F_{ss} - \Sigma_{ss,ds}) + \beta_D(\Sigma_{ss,SD} - \Sigma_{ss,DD})$$

Covarianza Abuela paterna – nieto:

$$\text{cov}(a_{ds}, a_x) = 0.5\Sigma_{ds,s} + 0.5\Sigma_{ds,D} - \beta_s(1 + F_{ds} - \Sigma_{ss,ds}) + \beta_D(\Sigma_{ds,SD} - \Sigma_{ds,DD})$$

Covarianza Abuelo materno – nieto:

$$\text{cov}(a_{sd}, a_x) = 0.5\Sigma_{sd,s} + 0.5\Sigma_{sd,D} + \beta_s(\Sigma_{sd,ss} - \Sigma_{sd,ds}) + \beta_D(1 + F_{SD} - \Sigma_{sd,DD})$$

Covarianza Abuela materna – nieto:

$$\text{cov}(a_{dd}, a_x) = 0.5\Sigma_{dd,s} + 0.5\Sigma_{dd,D} + \beta_s(\Sigma_{ss,DD} - \Sigma_{ds,DD}) - \beta_D(1 + F_{DD} - \Sigma_{dd,DD})$$

Empleando estas covarianzas, la matriz Σ_A en [2.5] es igual a:

$$\Sigma_A = \begin{bmatrix} (1 + F_{ss}) & \Sigma_{ss,ds} & \Sigma_{ss,sD} & \Sigma_{ss,DD} & \Sigma_{ss,s} & \Sigma_{ss,D} \\ & (1 + F_{ds}) & \Sigma_{ds,SD} & \Sigma_{ds,DD} & \Sigma_{ds,s} & \Sigma_{ds,D} \\ & & (1 + F_{sd}) & \Sigma_{sd,DD} & \Sigma_{sd,s} & \Sigma_{sd,D} \\ & & & (1 + F_{DD}) & \Sigma_{D,D,s} & \Sigma_{D,D,D} \\ & & & & (1 + F_s) & \Sigma_{s,D} \\ & & & & & (1 + F_D) \end{bmatrix} \quad [2.7]$$

simétrica

Expresiones matriciales para AR: Para un pedigrí con q individuos, el vector \mathbf{a} de orden $q \times 1$ está conformado por los VC, mientras que el vector ϕ (también de orden q) es el resultado de ordenar los residuos Mendelianos del mismo modo que en el vector \mathbf{a} .

Generalizando [2.5], $\text{Var}(\phi) = \mathbf{D} \sigma_A^2$ es una matriz diagonal cuyos elementos distintos de cero son iguales a $D_{i,i} = \text{Var}(\phi_i) / \sigma_A^2$.

Sea \mathbf{B} una matriz de orden $q \times q$ cuya fila i contiene los coeficientes de regresión parcial de abuelo paterno (β_S) y materno (β_D), los valores 0.5 para los padres, y ceros para los restantes animales, resultando en el animal i -ésimo igual a:

$$\mathbf{B}_i = [0 \quad \beta_{S_i} \quad -\beta_{S_i} \quad \dots 0 \quad \dots 0 \quad \beta_{D_i} \quad -\beta_{D_i} \quad 0.5 \quad 0.5 \quad 0],$$

Se puede escribir el sistema lineal recursivo [2.1] en forma vectorial-matricial

$$\mathbf{a} = \mathbf{B}\mathbf{a} + \phi \quad [2.8]$$

Trabajando esta última expresión se llega al sistema siguiente:

$$\mathbf{a} - \mathbf{B}\mathbf{a} = \phi \Rightarrow (\mathbf{I} - \mathbf{B})\mathbf{a} = \phi \quad [2.9]$$

Como la inversa de la matriz cuadrada $(\mathbf{I} - \mathbf{B})$ existe, la solución única para el sistema lineal a la derecha de [2.9] existe. Se puede probar la existencia de la inversa de $(\mathbf{I} - \mathbf{B})$ a partir del lema 2.1 en Berman y Plemmons (1994). En el texto citado, los autores prueban que el autovalor más grande de \mathbf{B} ($\rho(\mathbf{B})$) es tal que $\rho(\mathbf{B}) < 1$ si y solo si $(\mathbf{I} - \mathbf{B})^{-1}$ existe y es igual a:

$$(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{I} + \sum_{k=1}^{\infty} \mathbf{B}^k = \mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \mathbf{B}^3 + \dots, \quad k \geq 0 \quad [2.10]$$

Dejando que k se acerque al infinito. Por los tanto, la solución única del sistema a la derecha de [2.9] es igual a:

$$\mathbf{a} = (\mathbf{I} - \mathbf{B})^{-1} \phi \quad [2.11]$$

La esperanza de \mathbf{a} resulta de tomar el valor esperado en [2.11] para obtener:

$$\mathbf{E}(\mathbf{a}) = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{E}(\phi) = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{0} = \mathbf{0} \quad [2.12]$$

Mientras que la matriz de covarianzas entre los VC en [2.11] resulta igual a:

$$\begin{aligned}\text{Var}(\mathbf{a}) &= \text{Var}\left[(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\phi}\right] = (\mathbf{I} - \mathbf{B})^{-1} \text{Var}(\boldsymbol{\phi})(\mathbf{I} - \mathbf{B}')^{-1} \\ &= (\mathbf{I} - \mathbf{B})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{B}')^{-1} \sigma_A^2\end{aligned}\quad [2.13]$$

Con la estructura de covarianzas ($\boldsymbol{\Sigma}$) igual a:

$$\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{B}')^{-1}\quad [2.14]$$

matriz que es simétrica y positiva definida ($\boldsymbol{\Sigma} > 0$). Dado que los elementos diagonales de \mathbf{D} son varianzas o valores no-negativos, la inversa de $\boldsymbol{\Sigma}$ es:

$$\boldsymbol{\Sigma}^{-1} = (\mathbf{I} - \mathbf{B}') \mathbf{D}^{-1} (\mathbf{I} - \mathbf{B})\quad [2.15]$$

La estructura de covarianza de los VC en [2.14] y su inversa en [2.15] son generalizaciones de la matriz de relaciones aditivas \mathbf{A} y su inversa, respectivamente (Henderson, 1984; Quaas, 1988). Para comprender esta analogía, la i -ésima fila de la matriz \mathbf{P} es igual a

$$\mathbf{P}_i = [0 \quad 0 \quad 0 \quad \dots 0 \quad \dots 0 \quad 0 \quad 0 \quad 0.5 \quad 0.5 \quad 0]$$

Los valores 0.5 aparecen en las columnas correspondientes a los padres, mientras que los valores en las columnas de los abuelos son iguales a cero (compárese con la expresión de \mathbf{B}_i en la página anterior). El rol de la matriz \mathbf{P} en \mathbf{A} es equivalente al de la matriz \mathbf{B} en $\boldsymbol{\Sigma}$. Mientras que los únicos términos distintos de cero en las filas de \mathbf{P} , son los dos valores 0.5 en las columnas correspondientes a los padres. En el caso de \mathbf{B} , sus filas además contienen los coeficientes de regresión parcial β que conectan al VC de X con los de sus abuelos.

2.2. Estimación de los parámetros β

Cantet et al. (2017) definieron los coeficientes de regresión parcial paterno (β_s) y materno (β_D) como:

$$\beta_s = \frac{\Sigma_{\text{PGP},X} - \Sigma_{\text{PGP},R} \Sigma_{\text{RR}}^{-1} \Sigma_{\text{R},\text{PGP}}}{\Sigma_{\text{PGP},\text{PGP}} - \Sigma_{\text{PGP},R} \Sigma_{\text{RR}}^{-1} \Sigma_{\text{R},\text{PGP}}}\quad \beta_D = \frac{\Sigma_{\text{MGP},X} - \Sigma_{\text{MGP},R} \Sigma_{\text{RR}}^{-1} \Sigma_{\text{R},\text{MGP}}}{\Sigma_{\text{MGP},\text{MGP}} - \Sigma_{\text{MGP},R} \Sigma_{\text{RR}}^{-1} \Sigma_{\text{R},\text{MGP}}}\quad [2.16]$$

donde las funciones de los VC PGP y MGP se definen como $a_{\text{PGP}} = 0.5(a_{\text{PGS}} - a_{\text{PGD}})$ y $a_{\text{MGP}} = 0.5(a_{\text{MGS}} - a_{\text{MGD}})$. Todos los elementos de Σ son función de los β , pero la expresión para las covarianzas ancestro – X son distintas de la fórmula para obtener las covarianzas cuando ningún individuo del par es ancestro del otro (Cantet, 2019, comunicación personal).

Roverato (2000) obtuvo la distribución estadística de una matriz de covarianzas de variables aleatorias que siguen una distribución normal multivariada, donde la matriz inversa de la estructura de covarianzas puede ser descompuesta por el método sin raíces cuadradas de Cholesky tal que $\Sigma^{-1} = [(I - B') D^{-1/2}] [D^{-1/2} (I - B)]$. Dicho autor encontró que las distribuciones de los elementos diagonales de D , o varianzas de los residuos Mendelianos siguen distribuciones chi-cuadrado, mientras que los ubicadas fuera de las diagonales en la matriz triangular inferior $(I - B)$ son $N(0,1)$. Asimismo, Roverato (2000) observó que ambos grupos de variables aleatorias son independientes. Estas distribuciones permiten emplear distintos algoritmos de estimación de β_s y β_D para cada animal, aprovechando la independencia entre las filas de $(I - B)$ y entre ellas con los elementos de D .

A continuación, se discutirá cómo estimar β_s y β_D asumiendo que los cuatro abuelos y ambos padres se encuentran genotipados, así como el individuo focal X. Los dos estadísticos suficientes son $[\text{cov}(a_{\text{SS}}, a_x) - \text{cov}(a_{\text{DS}}, a_x)]$ y $[\text{cov}(a_{\text{DS}}, a_x) - \text{cov}(a_{\text{DD}}, a_x)]$ (Cantet et al, 2017). La ecuación de estimación relevante es $\Sigma_{\text{AX}} = \Sigma_A B_X$ cuya solución es $B_X = \Sigma_A^{-1} \Sigma_{\text{AX}}$. Sin embargo, los dos últimos elementos de B_X son iguales a 0.5 y no requieren ser estimados. Consecuentemente, se necesita transformar el sistema $\Sigma_{\text{AX}} = \Sigma_A B_X$ en uno reducido de orden 2, cuya solución contenga los parámetros, i.e. $\beta' = [\beta_s, \beta_D]$. Para ello se consideran las siguientes combinaciones lineales:

$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0.5 \frac{1}{\beta_s} & 0 \\ 0 & 0 & 1 & -1 & 0 & 0.5 \frac{1}{\beta_D} \end{bmatrix}$$

de modo tal que se pueda escribir $\mathbf{B}_x = \mathbf{L}' \boldsymbol{\beta}$. Entonces:

$$\boldsymbol{\Sigma}_A \mathbf{B}_x = \boldsymbol{\Sigma}_{AX} \Rightarrow \boldsymbol{\Sigma}_A \mathbf{L}' \boldsymbol{\beta} = \boldsymbol{\Sigma}_{AX} \Rightarrow \mathbf{L} \boldsymbol{\Sigma}_A \mathbf{L}' \boldsymbol{\beta} = \mathbf{L} \boldsymbol{\Sigma}_{AX} \quad [2.17]$$

La representación $\mathbf{L} \boldsymbol{\Sigma}_A \mathbf{L}' \boldsymbol{\beta} = \mathbf{L} \boldsymbol{\Sigma}_{AX}$, muestra cómo se puede reducir \mathbf{B}_x al vector $\boldsymbol{\beta}$.

Para desarrollar el sistema reducido se escribirán los elementos de $\boldsymbol{\Sigma}_A \mathbf{B}_x = \boldsymbol{\Sigma}_{AX}$ en más detalle como sigue:

$$\begin{bmatrix} (1 + F_{SS}) & \Sigma_{SS,DS} & \Sigma_{SS,SD} & \Sigma_{SS,DD} & \Sigma_{SS,S} & \Sigma_{SS,D} \\ & (1 + F_{DS}) & \Sigma_{DS,SD} & \Sigma_{DS,DD} & \Sigma_{DS,S} & \Sigma_{DS,D} \\ & & (1 + F_{SD}) & \Sigma_{SD,DD} & \Sigma_{SD,S} & \Sigma_{SD,D} \\ & & & (1 + F_{DD}) & \Sigma_{DD,S} & \Sigma_{DD,D} \\ & \text{simétrico} & & & (1 + F_S) & \Sigma_{S,D} \\ & & & & & (1 + F_D) \end{bmatrix} \begin{bmatrix} \beta_s \\ -\beta_s \\ \beta_D \\ -\beta_D \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} \Sigma_{SS,X} \\ \Sigma_{DS,X} \\ \Sigma_{SD,X} \\ \Sigma_{DD,X} \\ \Sigma_{S,X} \\ \Sigma_{D,X} \end{bmatrix} \quad [2.18]$$

La expresión [2.18] incluye las siguientes ecuaciones:

Ecuación 1:

$$(1 + F_{SS})\beta_s - \Sigma_{SS,DS}\beta_s + \Sigma_{SS,SD}\beta_D - \Sigma_{SS,DD}\beta_D + 0.5\Sigma_{SS,S} + 0.5\Sigma_{SS,D} = \Sigma_{SS,X}$$

$$(1 + F_{SS} - \Sigma_{SS,DS})\beta_s + (\Sigma_{SS,SD} - \Sigma_{SS,DD})\beta_D + 0.5(\Sigma_{SS,S} + \Sigma_{SS,D}) = \Sigma_{SS,X}$$

$$(1 + F_{SS} - \Sigma_{SS,DS})\beta_s + (\Sigma_{SS,SD} - \Sigma_{SS,DD})\beta_D = \Sigma_{SS,X} - 0.5(\Sigma_{SS,S} + \Sigma_{SS,D}) \quad \square$$

Ecuación 2:

$$\Sigma_{SS,DS}\beta_s - (1 + F_{DS})\beta_s + \Sigma_{DS,SD}\beta_D - \Sigma_{DS,DD}\beta_D + 0.5\Sigma_{DS,S} + 0.5\Sigma_{DS,D} = \Sigma_{DS,X}$$

$$(\Sigma_{SS,DS} - (1 + F_{DS}))\beta_s + (\Sigma_{DS,SD} - \Sigma_{DS,DD})\beta_D + 0.5(\Sigma_{DS,S} + \Sigma_{DS,D}) = \Sigma_{DS,X}$$

$$-(1 + F_{DS} - \Sigma_{SS,DS})\beta_s + (\Sigma_{DS,SD} - \Sigma_{DS,DD})\beta_D = \Sigma_{DS,X} - 0.5(\Sigma_{DS,S} + \Sigma_{DS,D}) \quad \square$$

Ecuación 3:

$$\begin{aligned} \Sigma_{SS,SD} \beta_s - \Sigma_{DS,SD} \beta_s + (1 + F_{SD}) \beta_D - \Sigma_{SD,DD} \beta_D + 0.5 \Sigma_{SD,S} + 0.5 \Sigma_{SD,D} &= \Sigma_{SD,X} \\ (\Sigma_{SS,SD} - \Sigma_{DS,SD}) \beta_s + ((1 + F_{SD}) - \Sigma_{SD,DD}) \beta_D + 0.5 (\Sigma_{SD,S} + \Sigma_{SD,D}) &= \Sigma_{SD,X} \\ (\Sigma_{SS,SD} - \Sigma_{DS,SD}) \beta_s + (1 + F_{SD} - \Sigma_{SD,DD}) \beta_D &= \Sigma_{SD,X} - 0.5 (\Sigma_{SD,S} + \Sigma_{SD,D}) \quad \square \end{aligned}$$

Ecuación 4:

$$\begin{aligned} \Sigma_{SS,DD} \beta_s - \Sigma_{DS,DD} \beta_s + \Sigma_{SD,DD} \beta_D - (1 + F_{DD}) \beta_D + 0.5 \Sigma_{DD,S} + 0.5 \Sigma_{DD,D} &= \Sigma_{DD,X} \\ (\Sigma_{SS,DD} - \Sigma_{DS,DD}) \beta_s + (\Sigma_{SD,DD} - (1 + F_{DD})) \beta_D + 0.5 (\Sigma_{DD,S} + \Sigma_{DD,D}) &= \Sigma_{DD,X} \\ (\Sigma_{SS,DD} - \Sigma_{DS,DD}) \beta_s - (1 + F_{DD} - \Sigma_{SD,DD}) \beta_D &= \Sigma_{DD,X} - 0.5 (\Sigma_{DD,S} + \Sigma_{DD,D}) \quad \square \end{aligned}$$

Ecuación 5:

$$\begin{aligned} \Sigma_{SS,S} \beta_s - \Sigma_{DS,S} \beta_s + \Sigma_{SD,S} \beta_D - \Sigma_{DD,S} \beta_D + 0.5(1 + F_s) + 0.5 \Sigma_{s,D} &= \Sigma_{s,X} \\ (\Sigma_{SS,S} - \Sigma_{DS,S}) \beta_s + (\Sigma_{SD,S} - \Sigma_{DD,S}) \beta_D + 0.5(1 + F_s + \Sigma_{s,D}) &= \Sigma_{s,X} \\ (\Sigma_{SS,S} - \Sigma_{DS,S}) \beta_s + (\Sigma_{SD,S} - \Sigma_{DD,S}) \beta_D &= \Sigma_{s,X} - 0.5(1 + F_s + \Sigma_{s,D}) \quad \square \end{aligned}$$

Ecuación 6:

$$\begin{aligned} \Sigma_{SS,D} \beta_s - \Sigma_{DS,D} \beta_s + \Sigma_{SD,D} \beta_D - \Sigma_{DD,D} \beta_D + 0.5 \Sigma_{s,D} + 0.5(1 + F_D) &= \Sigma_{D,X} \\ (\Sigma_{SS,D} - \Sigma_{DS,D}) \beta_s + (\Sigma_{SD,D} - \Sigma_{DD,D}) \beta_D + 0.5(1 + F_D + \Sigma_{s,D}) &= \Sigma_{D,X} \\ (\Sigma_{SS,D} - \Sigma_{DS,D}) \beta_s + (\Sigma_{SD,D} - \Sigma_{DD,D}) \beta_D &= \Sigma_{D,X} - 0.5(1 + F_D + \Sigma_{s,D}) \quad \square \end{aligned}$$

Las ecuaciones 1 a 6 se pueden trabajar para luego obtener:

$$\begin{bmatrix} (1+F_{SS}-\Sigma_{SS,DS}) & (\Sigma_{SS,SD}-\Sigma_{SS,DD}) \\ -(1+F_{DS}-\Sigma_{SS,DS}) & (\Sigma_{DS,SD}-\Sigma_{DS,DD}) \\ (\Sigma_{SS,SD}-\Sigma_{DS,SD}) & (1+F_{SD}-\Sigma_{SD,DD}) \\ (\Sigma_{SS,DD}-\Sigma_{DS,DD}) & -(1+F_{DD}-\Sigma_{SD,DD}) \\ (\Sigma_{SS,S}-\Sigma_{DS,S}) & (\Sigma_{SD,S}-\Sigma_{DD,S}) \\ (\Sigma_{SS,D}-\Sigma_{DS,D}) & (\Sigma_{SD,D}-\Sigma_{DD,D}) \end{bmatrix} \begin{bmatrix} \beta_S \\ \beta_D \end{bmatrix} = \begin{bmatrix} \Sigma_{SS,X}-0.5(\Sigma_{SS,S}+\Sigma_{SS,D}) \\ \Sigma_{DS,X}-0.5(\Sigma_{DS,S}+\Sigma_{DS,D}) \\ \Sigma_{SD,X}-0.5(\Sigma_{SD,S}+\Sigma_{SD,D}) \\ \Sigma_{DD,X}-0.5(\Sigma_{DD,S}+\Sigma_{DD,D}) \\ \Sigma_{S,X}-0.5(1+F_S+\Sigma_{S,D}) \\ \Sigma_{D,X}-0.5(1+F_D+\Sigma_{S,D}) \end{bmatrix} \quad [2.19]$$

La matriz de la derecha y el vector de la izquierda en [2.18] pueden escribirse como en un modelo lineal iguales a:

$$\mathbf{X} = \begin{bmatrix} (1+F_{SS}-\Sigma_{SS,DS}) & (\Sigma_{SS,SD}-\Sigma_{SS,DD}) \\ -(1+F_{DS}-\Sigma_{SS,DS}) & (\Sigma_{DS,SD}-\Sigma_{DS,DD}) \\ (\Sigma_{SS,SD}-\Sigma_{DS,SD}) & (1+F_{SD}-\Sigma_{SD,DD}) \\ (\Sigma_{SS,DD}-\Sigma_{DS,DD}) & -(1+F_{DD}-\Sigma_{SD,DD}) \\ (\Sigma_{SS,S}-\Sigma_{DS,S}) & (\Sigma_{SD,S}-\Sigma_{DD,S}) \\ (\Sigma_{SS,D}-\Sigma_{DS,D}) & (\Sigma_{SD,D}-\Sigma_{DD,D}) \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \Sigma_{SS,X}-0.5(\Sigma_{SS,S}+\Sigma_{SS,D}) \\ \Sigma_{DS,X}-0.5(\Sigma_{DS,S}+\Sigma_{DS,D}) \\ \Sigma_{SD,X}-0.5(\Sigma_{SD,S}+\Sigma_{SD,D}) \\ \Sigma_{DD,X}-0.5(\Sigma_{DD,S}+\Sigma_{DD,D}) \\ \Sigma_{S,X}-0.5(1+F_S+\Sigma_{S,D}) \\ \Sigma_{D,X}-0.5(1+F_D+\Sigma_{S,D}) \end{bmatrix} \quad [2.20]$$

Nótese que,

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{L}\Sigma_A\mathbf{L}')^{-1}\mathbf{L}\Sigma_A\mathbf{\beta} \quad [2.21]$$

$\mathbf{L}\Sigma_A\mathbf{L}'$ admite inversa debido a que $\Sigma_A > 0$ con todos sus elementos positivos.

Los elementos de la matriz de orden 2 en [2.21] son iguales a:

$$\begin{aligned} [\mathbf{X}'\mathbf{X}]_{11} &= (1+F_{SS}-\Sigma_{SS,DS})^2 + (1+F_{DS}-\Sigma_{SS,DS})^2 + (\Sigma_{SS,SD}-\Sigma_{DS,SD})^2 \\ &\quad + (\Sigma_{SS,DD}-\Sigma_{DS,DD})^2 + (\Sigma_{SS,S}-\Sigma_{DS,S})^2 + (\Sigma_{SS,D}-\Sigma_{DS,D})^2 \end{aligned} \quad [2.22]$$

$$\begin{aligned} [\mathbf{X}'\mathbf{X}]_{12} &= (1+F_{SS}-\Sigma_{SS,DS})(\Sigma_{SS,SD}-\Sigma_{SS,DD}) - (1+F_{DS}-\Sigma_{SS,DS})(\Sigma_{DS,SD}-\Sigma_{DS,DD}) \\ &\quad + (\Sigma_{SS,SD}-\Sigma_{DS,SD})(1+F_{SD}-\Sigma_{SD,DD}) - (\Sigma_{SS,DD}-\Sigma_{DS,DD})(1+F_{DD}-\Sigma_{SD,DD}) \\ &\quad + (\Sigma_{SS,S}-\Sigma_{DS,S})(\Sigma_{SD,S}-\Sigma_{DD,S}) + (\Sigma_{SS,D}-\Sigma_{DS,D})(\Sigma_{SD,D}-\Sigma_{DD,D}) \end{aligned} \quad [2.23]$$

$$\begin{aligned} [\mathbf{X}'\mathbf{X}]_{22} &= (\Sigma_{SS,SD}-\Sigma_{SS,DD})^2 + (\Sigma_{DS,SD}-\Sigma_{DS,DD})^2 + (1+F_{SD}-\Sigma_{SD,DD})^2 \\ &\quad + (1+F_{DD}-\Sigma_{SD,DD})^2 + (\Sigma_{SD,S}-\Sigma_{DD,S})^2 + (\Sigma_{SD,D}-\Sigma_{DD,D})^2 \end{aligned} \quad [2.24]$$

Mientras que los elementos del vector en [2.21] son iguales a:

$$\begin{aligned}
 [\mathbf{X}'\mathbf{y}]_1 = & \\
 & (1 + F_{SS} - \Sigma_{SS,DS}) (\Sigma_{SS,X} - 0.5(\Sigma_{SS,S} + \Sigma_{SS,D})) - \\
 & (1 + F_{DS} - \Sigma_{SS,DS}) (\Sigma_{DS,X} - 0.5(\Sigma_{DS,S} + \Sigma_{DS,D})) + \\
 & (\Sigma_{SS,SD} - \Sigma_{DS,SD}) (\Sigma_{SD,X} - 0.5(\Sigma_{SD,S} + \Sigma_{SD,D})) + \\
 & (\Sigma_{SS,DD} - \Sigma_{DS,DD}) (\Sigma_{DD,X} - 0.5(\Sigma_{DD,S} + \Sigma_{DD,D})) + \\
 & (\Sigma_{SS,S} - \Sigma_{DS,S}) (\Sigma_{S,X} - 0.5(1 + F_S + \Sigma_{S,D})) + \\
 & (\Sigma_{SS,D} - \Sigma_{DS,D}) (\Sigma_{D,X} - 0.5(1 + F_D + \Sigma_{S,D}))
 \end{aligned} \tag{2.25}$$

$$\begin{aligned}
 [\mathbf{X}'\mathbf{y}]_2 = & \\
 & (\Sigma_{SS,SD} - \Sigma_{SS,DD}) (\Sigma_{SS,X} - 0.5(\Sigma_{SS,S} + \Sigma_{SS,D})) + \\
 & (\Sigma_{DS,SD} - \Sigma_{DS,DD}) (\Sigma_{DS,X} - 0.5(\Sigma_{DS,S} + \Sigma_{DS,D})) + \\
 & (1 + F_{SD} - \Sigma_{SD,DD}) (\Sigma_{SD,X} - 0.5(\Sigma_{SD,S} + \Sigma_{SD,D})) - \\
 & (1 + F_{DD} - \Sigma_{SD,DD}) (\Sigma_{D,X} - 0.5(\Sigma_{DD,S} + \Sigma_{DD,D})) + \\
 & (\Sigma_{SD,S} - \Sigma_{DD,S}) (\Sigma_{S,X} - 0.5(1 + F_S + \Sigma_{S,D})) + \\
 & (\Sigma_{SD,D} - \Sigma_{DD,D}) (\Sigma_{D,X} - 0.5(1 + F_D + \Sigma_{S,D}))
 \end{aligned} \tag{2.26}$$

En consecuencia, es posible obtener los estimadores de β_s y β_d para cada animal de manera recursiva. Primero, es necesario estimar la fracción de IBD compartida entre animales genotipados, mediante métodos como los que se discuten en la sección siguiente. Luego, se deben ordenar los animales por fecha de nacimiento, para poder calcular la matriz de cada animal con los elementos entre los antecesores obtenidos previamente, y las relaciones estimadas entre el animal procesado y sus padres y abuelos según [2.25] y [2.26], para finalmente resolver la ecuación [2.21]. Una vez estimados β_s y β_d se puede calcular el resto de la fila de la matriz Σ de orden q empleando la expresión abajo y a la derecha de la descomposición de Bartlett [2.3].

2.3. Cálculo de la fracción de genoma compartido IBD

La estimación de las relaciones de la fracción de genoma compartido IBD se fundamenta en el grado de parecido entre parientes. Generalizando a Malécot (1948), Guo (1995) definió la fracción de genoma compartida IBD por dos individuos a aquella que proviene de segmentos de un antecesor común a ambos animales sin sufrir recombinación ni mutación (Figura 1).

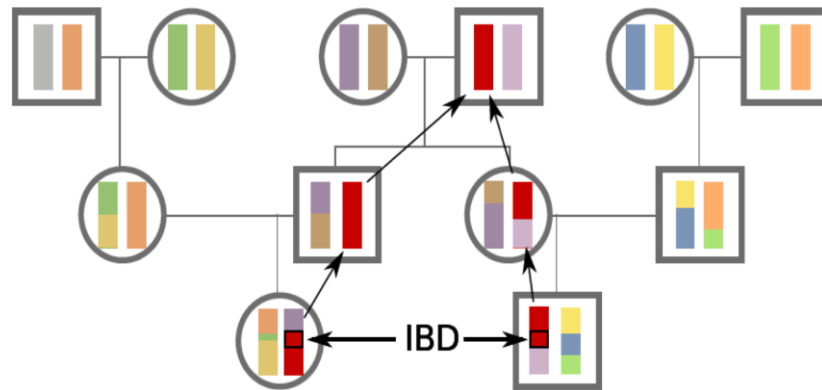


Figura 1. Diagrama de dos individuos que comparten segmentos idénticos por descendencia IBD, proveniente de un ancestro en común. Los cuadrados representan machos y los círculos hembras. Cada barra dentro de una figura representa la gameta paterna y materna.

La estimación de la fracción de IBD compartida entre dos animales requiere detectar aquellos segmentos de genoma iguales. Para ello, Donnelly (1983) desarrolló la noción de IBD “continuo”, a diferencia de la de IBD “discreto” (Malécot, 1948). Bickebøller y Thompson (1996) observaron que bajo la noción “discreta” de IBD no hay un número de genes equivalente en la noción “continua” (Cantet, R. J. C. 2019, comunicación personal). La dificultad es que en el caso discreto la noción de IBD aplica a “genes” que no son en general identificables individualmente, mientras que la noción continua apela a igualdad de bases o secuencias de ADN en todo el genoma. Por otra parte, para que cambie la probabilidad de IBD compartido $P(\text{IBD})$, las longitudes de los segmentos de ADN dependen del número de meiosis que separan a los dos individuos con el ancestro en común, del mismo modo que la $P(\text{IBD})$ discreta depende del número de generaciones entre cada animal en la relación estimada y el ancestro común. Así, segmentos IBD cortos indican una mayor distancia con el ancestro. Al mismo tiempo, la recombinación genera independencia entre los nuevos segmentos originados en la meiosis, pero como el número de recombinaciones por cromosoma rara vez es superior a 2 cada 100 cM, los estados

IBD de loci adyacentes no son generalmente independientes (Donnelly, 1983). Esta complejidad en el cálculo de las probabilidades IBD sobre todo el genoma, está sin embargo resuelta por algoritmos de cadenas de Markov Ocultas (“Hidden Markow Models”), condicional a la información de marcadores moleculares distribuidos a lo largo del genoma y del pedigrí, considerando el desequilibrio gamético existente en un panel denso de marcadores moleculares (SPN) y la consanguinidad. Este es el caso del algoritmo y programa desarrollados por Han y Abney (2011), cuya descripción escapa al presente marco teórico dada su complejidad.

Capítulo 3: Materiales y Métodos

Capítulo 3

Materiales y Métodos

3.1. Metodología

Para estimar los coeficientes de regresión parcial paterno (β_s), materno (β_D) y la consanguinidad (F_{AR}) se requiere estimar la matriz de covarianzas de orden 6 por “familia”, entre los cuatro abuelos y ambos padres del individuo. En esta tesis dicha matriz se calculó a partir de un arreglo triangular superior, compuesto por las relaciones de parentesco en la matriz de relaciones aditivas (A), mientras que el triángulo inferior y la diagonal contuvieron las relaciones de parentesco por IBD compartido observadas empleando información de un panel denso de marcadores SNP, las cuales de ahora en adelante serán denominadas RIBD. Para este cálculo se utilizó un algoritmo programado en Fortran 77 por R. J. C. Cantet.

A continuación, en la sección 3.2 se describe la simulación de RIBD bajo apareamiento aleatorio (RM, del inglés *Random Mating*) y selección (SEL) según describen Forneris (2016). En la sección 3.3 se presenta la estimación de las densidades de las RIBD abuelo nieto (RIBD-AN) junto con los parámetros de dichas distribuciones y los estimadores del núcleo o “kernel” de las distribuciones Beta y Normal. Luego, en la sección 3.4, se describen los métodos de estimación de las relaciones de parentesco observadas y realizadas. Se utilizó el método recursivo tabular para estimar A (Emik y Terril, 1949; Henderson, 1976), mientras que las relaciones de parentesco genómicas fueron estimadas mediante la fórmula propuesta por Forneris (2016), que introduce la noción de fracción de genoma IBD. En la sección 3.5, se describe el algoritmo de estimación de la matriz con RIBD familiares Σ de orden 6. A manera de ejemplo, se describe la derivación de la matriz de orden 3 entre ambos padres y su progenie común. La consideración de los efectos de la selección sobre las relaciones de parentesco y los parámetros se presenta en la sección 3.6, describiéndose allí la estimación de Σ condicional a los VC de los tatarabuelos y bisabuelos que permite realizar el ajuste por selección en Σ . Posteriormente, en la sección 3.7 se describe el coeficiente de consanguinidad bajo AR, Cantet (2019, comunicación personal). La sección 3.8 muestra la estructura del algoritmo computacional utilizado para la estimación de las RIBD

compartidas ancestro-individuo y de los parámetros del modelo AR. Finalmente, la sección 3.9 describe el análisis estadístico empleado para caracterizar las distribuciones.

3.2. Base de datos simulados

Se utilizaron dos conjuntos de datos simulados, uno bajo un sistema de apareamiento al azar (RM) ($N = 45095$) y el otro bajo selección (SEL) ($N = 46561$), en ambos casos replicados en cinco ocasiones, todas ellas generadas tal como se describe en Forneris et al. (2015). La simulación siguió los principios Mendelianos de herencia empleando el programa QMSim (Sargolzaei y Schenkel, 2009). La estructura de los datos fue la de un programa de mejoramiento genético porcino, formulado de manera simplificada con diez cromosomas (cinco pares autosómicos) de 160 cM de longitud y con 70000 SNPs. En una primera etapa se simularon las generaciones históricas para producir un nivel de desequilibrio gamético (LD) como el que se observa en una población real. Se distribuyeron aleatoriamente 35000 marcadores bialélicos a lo largo del genoma, recibiendo la posición de cada uno por muestreos en una distribución uniforme y con una frecuencia alélica igual a 0.5. Se alcanzó un equilibrio mutación-deriva en 2500 generaciones de apareamiento aleatorio con un tamaño efectivo igual a 500 y una tasa de mutación igual a 2×10^{-4} , seguido de un “cuello de botella” severo que redujo el tamaño efectivo poblacional a 75 durante 30 generaciones. En la segunda etapa, se generó la estructura poblacional más reciente resultado del proceso de selección. El sistema de apareamiento y selección aplicado durante cada una de cinco generaciones consistió en cruzar veinte verracos con doscientas cerdas en cada generación para minimizar la consanguinidad. Estos apareamientos resultaron en 2000 crías, siendo la mitad machos y la mitad hembras. De esa descendencia se seleccionaron los machos con mayores valores VC predichos por BLUP-modelo animal, mientras que las hembras se escogieron al azar. Los registros del pedigrí estuvieron disponibles para las cinco generaciones (10220 animales); así como la proporción realizable de genoma compartido entre cada par de animales.

3.3. Estimación de las densidades de las relaciones de IBD compartido abuelo - nieto y de los parámetros de la Regresión Ancestral

La contribución original de esta tesis es el haber obtenido distribuciones empíricas de las estimaciones de los parámetros de la Regresión Ancestral, de modo tal que se puedan ajustar β_s y β_d en bases de datos reales mediante algoritmos de verosimilitud o Bayesianos. Sean X_1, \dots, X_n un conjunto de variables aleatorias independientes e idénticamente distribuidas con una función de densidad f , la cual provee una descripción de la distribución de X con probabilidades asociadas iguales a $P(a < X < b) = \int_a^b f(x) dx$ para todo $a < b$ y pueda estimarse con un conjunto de muestras, supuestamente tomadas de f . Los estimadores del kernel de la densidad suelen ser métodos no paramétricos (Jones, 1990). Dado una muestra aleatoria X_1, \dots, X_n independientes e idénticamente distribuidas de una función de densidad f , el estimador del kernel (Silverman, 1998) es igual a:

$$\hat{f}(x) = (nh)^{-1} \sum_{i=1}^n K\{h^{-1}(x - X_i)\} \quad [3.1]$$

Los valores K y h son el kernel de la función y el parámetro de “ancho de banda” (o “suavidad”), respectivamente; n es el tamaño de la muestra y X_i es la i -ésimo variable aleatoria. Cuando el kernel es Gaussiano el estimador es igual a:

$$K(x) = \frac{1}{h\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2h^2}\right\} \text{ con } -\infty < x < \infty \quad [3.2]$$

El estimador $K(x)$ es una función integrable de valores reales no negativos y simétrica con respecto a μ , con lo cual se asegura que el promedio de la distribución correspondiente sea igual al de la muestra utilizada. La función estimada tiene las propiedades de continuidad y diferenciación de la función $K(x)$, de manera tal que si $K(x)$ es una función de densidad normal, la estimación de dicha función será una curva suave con derivadas en todo el espacio donde está definida la función.

La elección del parámetro de suavidad h es de crucial importancia para la estimación de la función de densidad y, al ser desconocido, debe estimarse. A tal efecto se suelen utilizar diferentes métodos que van desde una elección subjetiva (Silverman, 1998) hasta procedimientos más formales, como el propuesto por Jones et al. (1996). La elección con criterios subjetivos requiere graficar varias curvas y elegir la estimación más acorde con

la idea previa que se tiene de la forma de la densidad. Por otra parte, si se asume que $h \rightarrow 0$, el valor óptimo de h surge de la siguiente fórmula (Silverman, 1998):

$$h = \left[\frac{1}{2n\sqrt{\pi} \int (f''(x))^2} \right]^{\frac{1}{5}} \quad [3.3]$$

Para el estimador del kernel Gaussiano con varianza σ^2 con ϕ la densidad normal estándar,

$$\int f''(x)^2 dx = \sigma^{-5} \int \phi''(x)^2 dx = \frac{3}{8} \pi^{-1/2} \sigma^{-5} \approx 0.212 \sigma^{-5} \quad [3.4]$$

con lo cual, el parámetro de ancho de banda o suavidad es aproximado por

$$h = 1.06 \sigma n^{-\frac{1}{5}} \quad [3.5]$$

La esperanza del estimador del kernel [3.1] (Silverman, 1998) es igual a:

$$E(\hat{f}(x)) = E\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\right] = \frac{1}{h} E\left[K\left(\frac{x-X_i}{h}\right)\right] = \frac{1}{h} \int K\left(\frac{x-X_i}{h}\right) f(x) dx \quad [3.6]$$

Con respecto a la varianza:

$$\text{Var}(\hat{f}(x)) = E\left\{\left[\hat{f}(x) - E(\hat{f}(x))\right]^2\right\} = E(\hat{f}(x)^2) - \left(E(\hat{f}(x))\right)^2$$

Con lo cual, empleando [3.1], se obtiene:

$$\text{Var}(\hat{f}(x)) = \frac{1}{nh^2} \int K\left(\frac{x-X_i}{h}\right)^2 f(x) dx - \left\{\frac{1}{h} \int K\left(\frac{x-X_i}{h}\right) f(x) dx\right\}^2 \quad [3.7]$$

Estimador kernel Beta: El estimador kernel de la densidad de una variable con distribución Beta resulta igual a la expresión siguiente (Chen, 1999):

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x/(b+1), (1-x)/(b+1)}(X_i) \quad [3.8]$$

donde,

$$K(u; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} u^{\alpha-1} (1-u)^{\beta-1}, u \in [0, 1] \quad [3.9]$$

La constante en [3.9] es definida en términos de una función gamma

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

La esperanza y la varianza (Chen, 1999) son iguales a:

$$E\{\hat{f}_n(x)\} = \int_0^1 \sum_{i=1}^n K_{x/(b+1), (1-x)/(b+1)}(y) f(y) dy = E\{f(\xi_x)\} \quad [3.10]$$

Donde ξ_x es igual a una variable aleatoria $\sim Beta\left\{\frac{x}{\beta+1}, \frac{1-x}{\beta+1}\right\}$ cuya varianza, es igual a:

$$\begin{aligned} \text{Var}\{\hat{f}_n(x)\} &= \frac{\text{Var}\{K(x_i, x/b+1, (1-x)/b+1)\}}{n} \\ &= \frac{\left[E\{K(x_i, x/b+1, (1-x)/b+1)\}^2 - \left[E\{\hat{f}_b(x)\} \right]^2 \right]}{n} \end{aligned} \quad [3.11]$$

Momentos de la distribución Beta:

Las propiedades de la distribución Beta son las siguientes:

i) La constante de la distribución es simétrica en sus parámetros, es decir $B(\alpha, \beta) = B(\beta, \alpha)$,

$$ii) B(\alpha, 1) = \frac{1}{\alpha},$$

$$iii) B(\alpha + 1, \beta) = \frac{\alpha}{\alpha + \beta} B(\alpha, \beta).$$

Sea, $X \sim Beta(\alpha, \beta)$ con $\alpha > 0$ y $\beta > 0$, su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{si } x \in (0, 1). \\ 0, & \text{para cualquier otro valor de } x. \end{cases},$$

1er Momento:

$$\begin{aligned}
 E(x) &= \int_{-\infty}^{\infty} x f(x) dx \\
 &= \int_0^1 x \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{(\alpha+1)-1} (1-x)^{\beta-1} dx \\
 &= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\
 &= \frac{\frac{\alpha}{\alpha+\beta} B(\alpha, \beta)}{B(\alpha, \beta)} \quad \text{por la propiedad iii.} \\
 &= \frac{\alpha}{\alpha+\beta} \quad \blacksquare \tag{3.12}
 \end{aligned}$$

2do Momento:

$$\begin{aligned}
 E(x^2) &= \int_0^1 x^2 \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\
 &= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{(\alpha+2)-1} (1-x)^{\beta-1} dx \\
 &= \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} \\
 &= \frac{B((\alpha+1)+1, \beta)}{B(\alpha, \beta)} \\
 &= \frac{\alpha+1}{\alpha+1+\beta} \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \quad \text{por la propiedad iii.} \\
 &= \frac{\alpha+1}{\alpha+1+\beta} \frac{\alpha}{\alpha+\beta} \quad \blacksquare
 \end{aligned}$$

Finalmente, la varianza de la variable X es,

$$\begin{aligned}
 \text{Var}(x) &= E(x^2) - E^2(x) \\
 &= \frac{(\alpha + 1)\alpha}{(\alpha + 1 + \beta)(\alpha + \beta)} - \left(\frac{\alpha}{\alpha + \beta}\right)^2 \\
 &= \frac{1}{(\alpha + \beta + 1)(\alpha + \beta)^2} [(\alpha + 1)\alpha(\alpha + \beta) - \alpha^2(\alpha + 1 + \beta)] \\
 &= \frac{1}{(\alpha + \beta + 1)(\alpha + \beta)^2} [\alpha^3 + \alpha^2\beta + \alpha^2 + \alpha\beta - \alpha^3 - \alpha^2 - \alpha^2\beta] \\
 &= \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2} \quad \blacksquare \quad [3.13]
 \end{aligned}$$

Los estimadores de kernel fueron calculados con los datos simulados empleando PROC KDE de SAS.

3.4. Relaciones de parentesco

3.4.1. Relaciones esperadas

El ij -ésimo elemento fuera de la diagonal de la matriz de relaciones A es igual al numerador del coeficiente de parentesco (Wright, 1922) entre los individuos i y j , y el elemento diagonal ii tiene valor uno más el coeficiente de consanguinidad del individuo i (Wright, 1922; Quaas, 1976). La matriz A fue calculada recursivamente mediante el método tabular (Emik y Terril, 1949). Los individuos en A deben identificarse cronológicamente, de manera tal, que los padres precedan a su progenie, siendo los primeros b individuos los pertenecientes a la generación “base”, los cuales se asumen no ser consanguíneos ni estar emparentados entre sí. El cálculo determinístico de A fue realizado con el algoritmo de Henderson (1976).

3.4.2. Relaciones realizadas

Directamente se utilizaron las relaciones genómicas simuladas por Forneris (2016), sección 4.3.2, capítulo 4. La “verdadera” proporción del genoma compartido IBD por un par de individuos (g_T) proviene de una mezcla de distribuciones de densidad igual a:

$$f(g_T) = \sum_{k=1}^K \eta_k p(g_{T_k}) \quad [3.14]$$

En [3.14] $p(g_{T_k})$ representa la función de densidad de probabilidad condicional de g_T dada la relación de parentesco genómico de la clase k ($k = 1, \dots, K$ donde K es la cantidad de relaciones de parentesco utilizadas) (Figura 4.1 sección 4.4.1, capítulo 4 de Forneris (2016)) y η_k es el coeficiente de mixtura para la clase. Para los datos simulados, g_T se calculó en cada par de individuos, de la siguiente forma:

$$g_T = \frac{\sum_{c=1}^c l_c g_{T,c}}{L} \text{ y } L = \sum_{c=1}^c l_c \quad [3.14.A]$$

donde $g_{T,c}$, corresponde al valor de la proporción del genoma compartido en el cromosoma c de longitud l_c (Guo, 1995).

3.5. Estimación de la matriz de covarianzas Σ

Para el i -ésimo individuo la matriz cuadrada y simétrica Σ contiene las RIBD de abuelos y padres (orden 6), pero es “ajustada” por todas las recombinaciones ancestrales (tatarabuelos y bisabuelos) ocurridas en el pedigrí. A modo de ilustración se calcula Σ para el caso de padres e hijo (orden 3).

3.5.1. Matriz Σ de orden 3

La estructura de RIBD en Σ de orden 3 entre el padre (S), la madre (D) y el hijo (X) es igual a:

$$\Sigma = \begin{bmatrix} 1 + F_s & \Sigma_{SD} & \Sigma_{SX} \\ \Sigma_{SD} & 1 + F_D & \Sigma_{DX} \\ \Sigma_{SX} & \Sigma_{DX} & 1 + F_x \end{bmatrix} \quad [3.15]$$

Nótese que ambos padres representan aquí la población base, pero el algoritmo es general en el sentido que ambos progenitores pueden ser consanguíneos y, además, estar

emparentados. Las diagonales corresponden a uno más la consanguinidad del padre, la madre y el hijo, respectivamente, mientras que los elementos no diagonales son las RIBD tomadas de a pares entre los tres individuos.

Dado que Σ es simétrica y positiva definida admite una descomposición de Cholesky

$$\text{igual a } \Sigma = CC', \text{ con } C = \begin{bmatrix} S_{11} & 0 & 0 \\ S_{12} & S_{22} & 0 \\ S_{13} & S_{23} & S_{33} \end{bmatrix} \text{ y } C' = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ 0 & S_{22} & S_{23} \\ 0 & 0 & S_{33} \end{bmatrix}. \text{ Los elementos de } C$$

surgen de las siguientes expresiones escalares:

$$S_{11}^2 = \Sigma_{11} \quad \rightarrow \quad S_{11} = \sqrt{\Sigma_{11}} = \sqrt{1 + F_s}.$$

$$S_{12}S_{11} = \Sigma_{12} \quad \rightarrow \quad S_{12} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}} = \frac{\Sigma_{SD}}{\sqrt{1 + F_s}}.$$

$$S_{12}^2 + S_{22}^2 = \Sigma_{22} \quad \rightarrow \quad S_{22}^2 = \Sigma_{22} - S_{12}^2 \quad \rightarrow \quad S_{22} = \sqrt{\Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}} = \sqrt{1 + F_D - \frac{\Sigma_{SD}^2}{1 + F_s}}.$$

$$S_{13}S_{11} = \Sigma_{13} \quad \rightarrow \quad S_{13} = \frac{\Sigma_{13}}{\sqrt{\Sigma_{11}}} = \frac{\Sigma_{SX}}{\sqrt{1 + F_s}}.$$

$$S_{13}S_{12} + S_{23}S_{22} = \Sigma_{23} \quad \rightarrow \quad S_{23} = \frac{\Sigma_{23} - S_{13}S_{12}}{S_{22}} = \frac{\Sigma_{23} - \left[\frac{\Sigma_{13}\Sigma_{12}}{\Sigma_{11}} \right]}{\sqrt{\Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}}} = \frac{\Sigma_{DX} - \left[\frac{\Sigma_{SX}\Sigma_{SD}}{(1 + F_s)} \right]}{\sqrt{1 + F_D - \left[\frac{\Sigma_{SD}^2}{1 + F_s} \right]}}.$$

Finalmente,

$$S_{13}^2 + S_{23}^2 + S_{33}^2 = \Sigma_{33} \quad \rightarrow \quad S_{33}^2 = \Sigma_{33} - (S_{13}^2 + S_{23}^2)$$

$$\rightarrow \quad S_{33} = \sqrt{\Sigma_{33} - \frac{\Sigma_{13}^2}{\Sigma_{11}} - \frac{\Sigma_{11}\Sigma_{23}^2 + \Sigma_{13}^2\Sigma_{12}^2 - 2\Sigma_{11}\Sigma_{12}\Sigma_{13}\Sigma_{23}}{\Sigma_{11}(\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2)}}$$

$$\rightarrow \quad S_{33} = \sqrt{(1 + F_x) - \frac{\Sigma_{SX}^2}{1 + F_s} - \frac{(1 + F_s)^2 \Sigma_{DX}^2 + \Sigma_{SX}^2 \Sigma_{SD}^2 - 2(1 + F_s) \Sigma_{SD} \Sigma_{SX} \Sigma_{DX}}{(1 + F_s)((1 + F_s)(1 + F_D) - \Sigma_{SD}^2)}}.$$

Alternativamente, para mejorar el cálculo de operaciones de raíces cuadradas, la descomposición de Cholesky libre de raíces es igual a $\Sigma = \mathbf{H}\mathbf{D}\mathbf{H}' = \mathbf{H}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{H}' = \mathbf{C}\mathbf{C}'$, donde:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{H}_{12} & 1 & 0 \\ \mathbf{H}_{13} & \mathbf{H}_{23} & 1 \end{bmatrix} \quad \mathbf{D}^{1/2} = \begin{bmatrix} \mathbf{D}_1^{1/2} & 0 & 0 \\ 0 & \mathbf{D}_2^{1/2} & 0 \\ 0 & 0 & \mathbf{D}_3^{1/2} \end{bmatrix}$$

Los elementos de \mathbf{H} y \mathbf{D} se muestran en las siguientes expresiones escalares.

Elementos de D:

$$\mathbf{S}_{11} = \sqrt{\Sigma_{11}} = \mathbf{D}_1^{1/2} \quad \rightarrow \quad \mathbf{D}_1 = \Sigma_{11}.$$

$$\mathbf{D}_2^{1/2} = \mathbf{S}_{22} \quad \rightarrow \quad \mathbf{D}_2 = \mathbf{S}_{22}^2 = \Sigma_{22} - \frac{\Sigma_{12}^2}{\Sigma_{11}}.$$

$$\mathbf{S}_{33} = \mathbf{D}_3^{1/2} \quad \rightarrow \quad \mathbf{D}_3 = \mathbf{S}_{33}^2 = (1 + F_x) - \frac{\Sigma_{sx}^2}{(1 + F_s)} - \frac{(1 + F_s)^2 \Sigma_{dx}^2 + \Sigma_{sx}^2 \Sigma_{sd}^2 - 2(1 + F_s) \Sigma_{sd} \Sigma_{sx} \Sigma_{dx}}{(1 + F_s)^2 ((1 + F_s)(1 + F_d) - \Sigma_{sd}^2)}.$$

Elementos de H:

$$\mathbf{S}_{12} = \mathbf{H}_{12} \mathbf{D}_1^{1/2} \quad \rightarrow \quad \mathbf{H}_{12} = \mathbf{S}_{12} \mathbf{D}_1^{-1/2} = \frac{\Sigma_{12}}{\sqrt{\Sigma_{11}}} (\sqrt{\Sigma_{11}})^{-1} \quad \rightarrow \quad \mathbf{H}_{12} = \frac{\Sigma_{12}}{\Sigma_{11}}.$$

$$\mathbf{S}_{13} = \mathbf{H}_{13} \mathbf{D}_1^{1/2} \quad \rightarrow \quad \mathbf{H}_{13} = \mathbf{S}_{13} \mathbf{D}_1^{-1/2} = \frac{\Sigma_{13}}{\sqrt{\Sigma_{11}}} (\sqrt{\Sigma_{11}})^{-1} = \frac{\Sigma_{13}}{\Sigma_{11}}.$$

$$\mathbf{S}_{23} = \mathbf{H}_{23} \mathbf{D}_2^{1/2} \quad \rightarrow \quad \mathbf{H}_{23} = \mathbf{S}_{23} \mathbf{D}_2^{-1/2} = \frac{\Sigma_{dx} - \frac{\Sigma_{sx} \Sigma_{sd}}{1 + F_s}}{\sqrt{1 + F_d - \left[\frac{\Sigma_{sd}^2}{1 + F_s} \right]}} = \frac{(1 + F_s) \Sigma_{dx} - \Sigma_{sx} \Sigma_{sd}}{((1 + F_d)(1 + F_s) - \Sigma_{sd}^2) \sqrt{1 + F_s}}.$$

En términos de la matriz Σ los elementos de \mathbf{D} y \mathbf{H} son los siguientes:

$$\mathbf{D}_1 = \Sigma_{11} = 1 + F_s, \quad \mathbf{D}_2 = 1 + F_d - \frac{\Sigma_{sd}^2}{1 + F_s},$$

$$\mathbf{D}_3 = (1 + F_x) - \frac{\Sigma_{sx}^2}{(1 + F_s)} - \frac{(1 + F_s)^2 \Sigma_{dx}^2 + \Sigma_{sx}^2 \Sigma_{sd}^2 - 2(1 + F_s) \Sigma_{sd} \Sigma_{sx} \Sigma_{dx}}{(1 + F_s)((1 + F_s)(1 + F_d) - \Sigma_{sd}^2)},$$

$$H_{12} = \frac{\Sigma_{SD}}{(1+F_S)}, \quad H_{13} = \frac{\Sigma_{SX}}{(1+F_S)}, \quad H_{23} = \frac{(1+F_S)\Sigma_{DX} - \Sigma_{SX}\Sigma_{SD}}{\left(\left(1+F_D\right)\left(1+F_S\right) - \Sigma_{SD}^2\right)\sqrt{1+F_S}}.$$

3.6. Consideración de los efectos de la selección

Durante un proceso de selección las recombinaciones ancestrales deben ser tenidas en cuenta en el cálculo de Σ para evitar posibles sesgos debidos a la selección (Im et al., 1989). Por lo tanto, el algoritmo incluyó todos los VC ancestrales hasta la población base, lo que requirió considerar 30 meiosis en 16 tatarabuelos, 8 bisabuelos, 4 abuelos y 2 padres (Figura 2).

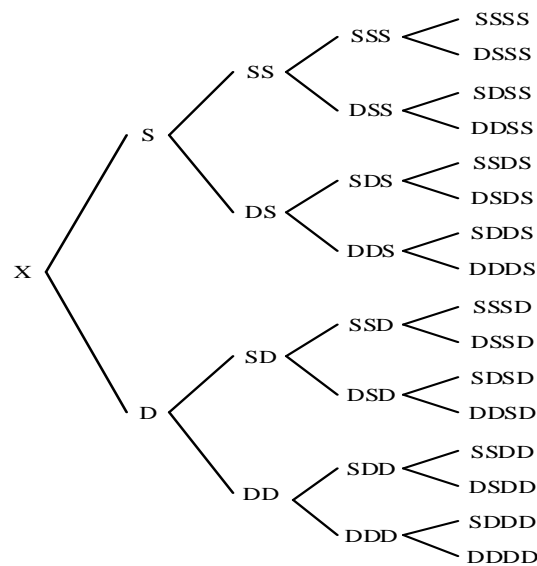


Figura 2. Pedigrí incluyendo hasta los tatarabuelos del individuo X. Los distintos nodos representan a sus padres (2 meiosis), los abuelos paternos y maternos (4 meiosis), los bisabuelos paternos y maternos (8 meiosis) y finalmente, los tatarabuelos paternos y maternos (16 meiosis).

Cuando se condiciona en las recombinaciones ocurridas entre parejas de bisabuelos y de tatarabuelos (Figura 2), la ecuación estructural autoregresiva del VC en los VC ancestrales, tiene un residuo Mendeliano independiente de todos los otros en la población, y la predicción AR no es multicolineal. Más formalmente, sea:

$$\Sigma_S = \begin{bmatrix} (\Sigma_{TT})_{24 \times 24} & (\Sigma_{TA})_{24 \times 6} \\ (\Sigma_{AT})_{6 \times 24} & (\Sigma_{AA})_{6 \times 6} \end{bmatrix}$$

La matriz RIBD entre los ancestros es Σ_{TT} y posee las relaciones IBD entre los bisabuelos y tatarabuelos; Σ_{TA} tiene en cuenta las RIBD entre estos ancestros y los cuatro abuelos junto a ambos padres. Asimismo, Σ_{AA} es la matriz de relaciones entre los cuatro abuelos y ambos padres. El complemento Schur (Harville, 2008) del bloque de interés Σ_{AA} de Σ_s es igual a:

$$\Sigma_{AA}^* = \Sigma_{AA} - \Sigma_{AT} \Sigma_{TT}^{-1} \Sigma_{TA} \quad [3.16]$$

La matriz Σ_{AA}^* considera el efecto de la selección, al incluir los VC de los ancestros (tatarabuelos y bisabuelos) selectos y refleja la independencia condicional con el resto del pedigrí, de modo similar al “sistema de transferencia” entre VC de Tallis (1987). Por último, el sistema de ecuaciones para estimar los parámetros de la Regresión Ancestral (β_s , β_D y F_{AR}) surge de sustituir la matriz Σ_{AA} por Σ_{AA}^* :

$$\begin{bmatrix} \Sigma_{AA}^* & \Sigma_{AA,X} \\ \Sigma_{X,AA} & \Sigma_{XX} \end{bmatrix} \begin{bmatrix} \beta_s \\ -\beta_s \\ \beta_D \\ -\beta_D \\ 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} \Sigma_{SS,X} \\ \Sigma_{DS,X} \\ \Sigma_{SD,X} \\ \Sigma_{DD,X} \\ \Sigma_{S,X} \\ \Sigma_{D,X} \end{bmatrix} \quad [3.17]$$

3.7. Coeficiente de endogamia (F_{AR})

Cantet (2019, comunicación personal) obtuvo la siguiente expresión para el coeficiente de consanguinidad bajo AR [2.4]:

$$F_{X(AR)} = \frac{1}{2} \left[\Sigma_{S,D} + \beta_s (\Sigma_{SS,D} - \Sigma_{DS,D}) + \beta_D (\Sigma_{S,SD} - \Sigma_{S,DD}) + \beta_s \beta_D (\Sigma_{SS,SD} - \Sigma_{SS,DD} - \Sigma_{DS,SD} + \Sigma_{DS,DD}) \right] \quad [3.18]$$

3.8. Estimación de los parámetros β_S , β_D y F_{AR}

Las relaciones de IBD compartido abuelo nieto (RIBD-AN), los parámetros β_S , β_D y F_{AR} , se estimaron mediante el siguiente algoritmo:

Algoritmo. Estimación de parámetros AR con información completa

Input: archivo genealógico (*ped*)

Input: archivo relaciones genómicas (*geno*)

- 1 Identificar individuos de *geno* en *ped*;
 - 2 Estimar F_{GENO} y F_{PED} por cada individuo;
 - 3 Identificar SS, DS, SD, DD;
 - 4 Construir Σ de orden 6 priorizando *geno* sobre *ped*;
 - 5 Construir A para cada individuo;
 - 6 Obtener Σ condicional a los VC de ancestros hasta la meiosis 30;
 - 7 Calcular RIBD-AN;
 - 8 Generar el sistema reducido y obtener la solución;
 - 9 Actualizar RIBD-AN y estimar F_{AR}
-

3.9. Análisis estadístico

Para analizar y comparar las estimaciones de los coeficientes de regresión parcial en los datos simulados, según la población sea selecta (SEL) o se aparee aleatoriamente (RM), se efectuó primero la prueba de Levene, buscando evidencias de heterogeneidad de varianza por réplica. Como consecuencia, se ajustó un modelo de una vía tomando como tratamiento RM o SEL y considerando heteroscedasticidad mediante una estructura de covarianzas diagonal. El análisis estadístico fue realizado en PROC MIXED de SAS, utilizando la corrección de los grados de libertad de Kenward y Roger (1997). Por último, se realizaron las pruebas de comparaciones múltiples entre tratamientos empleando el procedimiento de Tukey. En todos los casos se consideraron significativos los valores del estadístico asociados con $\alpha = 0.05$.

Capítulo 4: Resultados

Capítulo 4

Resultados

4.1. Distribución de las Relaciones de IBD compartido

El Cuadro 1 presenta la verosimilitud para las covarianzas estandarizadas o relaciones de IBD compartido (Guo, 1995) entre los distintos abuelos y el nieto (RIBD-AN), estimadas bajo el modelo de Regresión Ancestral. Para este conjunto de datos se calculó el logaritmo de la función de verosimilitud respecto de cada una de las siguientes distribuciones: Beta, Normal, Log-Normal, Weibull, Gamma y Exponencial, evaluadas en simulaciones producidas mediante apareamiento aleatorio (RM) o bajo selección (SEL). La segunda columna corresponde al tipo de RIBD-AN, siendo SS la covarianza con el abuelo paterno, DS con la abuela paterna, SD con el abuelo materno y DD con la abuela materna. Como criterio de selección se estableció al valor máximo del logaritmo de la función de verosimilitud.

Cuadro 1. Verosimilitudes de distribuciones ajustadas por Máxima Verosimilitud, para las RIBD-AN, estimadas bajo AR sobre RM y SEL.

Datos	Tipo	N	Distribuciones					
			Beta	Gamma	Weibull	Normal	Log.Normal	Exponencial
RM	SS	11 237	12 575.76 ^b	12 554.18	12 341.33	12 335.56	12 460.23	3 402.55
SEL		11 684	10 081.86 ^a	10 028.22	9 918.97	9 884.72	9 830.75	1 253.60
RM	DS	11 309	12 617.68 ^b	12 589.99	12 395.52	12 381.58	12 490.61	3 403.04
SEL		11 606	10 651.40 ^a	10 640.69	10 430.18	10 415.75	10 495.35	1 752.94
RM	SD	11 261	12 485.00 ^b	12 452.44	12 273.71	12 257.81	12 344.93	3 323.23
SEL		11 699	10 216.65 ^a	10 149.81	10 057.79	10 029.05	9 947.24	1 212.28
RM	DD	11 288	12 484.96 ^b	12 252.03	12 252.03	12 233.35	12 370.80	3 411.83
SEL		11 584	10 626.40 ^a	10 597.16	10 435.85	10 409.67	10 434.05	1 809.15

Valores con letras distintas entre RIBD para RM o SEL son significativamente diferentes. SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna). SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

En el Cuadro 1 se observa que la distribución Beta mostró los máximos valores de la función de verosimilitud con respecto a las otras distribuciones ajustadas, tanto para los datos RM como bajo SEL. Los parámetros de la distribución Beta mostraron la mayor

probabilidad de ocurrencia; los valores obtenidos fueron estadísticamente significativos ($P < 0.05$).

4.1.1. Parámetros de la distribución Beta

El Cuadro 2 muestra la estimación de los parámetros de la distribución Beta (alfa, α y beta, β), con sus respectivos errores estándar (EE) para las distintas RIBD-AN.

Cuadro 2. Parámetros de la distribución Beta con sus respectivos errores estándar (EE), para las RIBD-AN, estimadas bajo AR, sobre RM y SEL.

Datos	Tipo	Parámetros y Error Estándar			
		α	EE	β	EE
RM	SS	8.1193	0.1063	21.7502	0.2903
SEL		6.5292	0.0835	13.2210	0.1724
RM	DS	8.0863	0.1055	21.6071	0.2875
SEL		6.8407	0.0879	14.7721	0.1935
RM	SD	8.0468	0.1052	21.3313	0.2844
SEL		6.7165	0.0859	13.5289	0.1763
RM	DD	7.9033	0.1032	21.1550	0.2817
SEL		6.7753	0.0871	14.7448	0.1934

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna). SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

Del Cuadro 2 se infiere que la magnitud de α fue siempre menor que la de β , sugiriendo que en estos datos las distribuciones de las RIBD-AN fueron asimétricas hacia la derecha. Al examinar los errores estándar de estimación de α y β , no se encontró una gran discrepancia respecto de los valores estimados, los cuales rara vez superaron 0.20. El estadístico de Kolmogórov-Smirnov (D) para las RIBD-AN en ambos grupos de datos, es igual a la mayor diferencia absoluta entre la frecuencia acumulada observada $\hat{F}_n(x)$ y la teórica $F_0(x)$ obtenida de la distribución Beta bajo hipótesis nula. Tomando como valor crítico al cociente $1.95/\sqrt{n}$ y $\alpha = 0.01$, se concluye que no hay evidencias para rechazar la hipótesis nula y, por lo tanto, la distribución Beta muestra un buen ajuste a estos datos.

4.1.2. Estimadores de los momentos de la distribución Beta para RIBD-AN

En [3.12] del capítulo 3 se observa que el valor esperado de una distribución Beta es

$E(x) = \frac{\alpha}{\alpha + \beta}$, mientras que de [3.13] se infiere que la varianza es igual a

$\text{Var}(x) = \frac{\alpha \beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$. El Cuadro 3 presenta los estimadores de medias, varianzas,

desvíos estándar y coeficientes de variación de las distribuciones empíricas de las RIBD-AN, según se asocien con SS, DS, SD, DD, para datos bajo RM o SEL.

Cuadro 3. Estimadores de las esperanzas $E(X)$, varianzas $\text{Var}(X)$, medias muestrales \bar{X}_n , varianzas muestrales S_x^2 , desvíos estándar DS, y coeficientes de variación CV, para las RIBD-AN estimadas bajo AR bajo RM o SEL.

Datos	Tipo	$E(X)$	$\text{Var}(X)$	DS	CV	\bar{X}_n	S_x^2	DS
RM	SS	0.2718	0.0064	0.0801	0.2946	0.2718	0.0065	0.0807
SEL		0.3306	0.0107	0.1033	0.3124	0.3305	0.0108	0.1038
RM	DS	0.2723	0.0065	0.0803	0.2949	0.2723	0.0066	0.0810
SEL		0.3165	0.0096	0.0978	0.3090	0.3163	0.0097	0.0986
RM	SD	0.2739	0.0065	0.0809	0.2954	0.2739	0.0066	0.0815
SEL		0.3318	0.0104	0.1022	0.3079	0.3317	0.0105	0.1027
RM	DD	0.2720	0.0066	0.0812	0.2984	0.2719	0.0067	0.0819
SEL		0.3148	0.0096	0.0979	0.3109	0.3147	0.0097	0.0985

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna). SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

Se observa en el Cuadro 3 que las cuatro RIBD-AN provenientes de datos SEL fueron 1.22, 1.16, 1.21 y 1.16 veces mayores que aquellas observadas bajo RM, para SS, DS, SD y DD, respectivamente. Los valores de las medias observadas fueron muy similares a las esperadas, sugiriendo que las primeras fueron estimadores insesgados de las últimas. Asimismo, las RIBD-AN en las distribuciones SEL mostraron una leve dispersión y mayor CV, respecto de las relaciones bajo RM.

4.1.3. Curtosis, asimetría y otros estadísticos de las distribuciones empíricas

El Cuadro 4 presenta el mínimo, el máximo, la mediana, la media, el desvío estándar, la asimetría y la curtosis de las distribuciones empíricas para las RIBD-AN según SS, DS, SD o DD, calculadas para RM y SEL.

Cuadro 4. Mediana, media, desvío estándar, mínimo, máximo, asimetría y curtosis para las RIBD-AN, estimadas bajo AR sobre RM y SEL.

Datos	Tipo	Mediana	Media	DS	Mín.	Máx.	Asimetría	Curtosis
RM	SS	0.2672	0.2718	0.0807	0.125	0.5483	0.3450	2.5409
SEL		0.3261	0.3305	0.1038	0.125	0.792	0.3348	2.8381
RM	DS	0.2669	0.2723	0.0810	0.125	0.5545	0.3283	2.4976
SEL		0.3095	0.3163	0.0986	0.125	0.8391	0.4021	2.9135
RM	SD	0.2689	0.2739	0.0815	0.125	0.5668	0.3191	2.4961
SEL		0.3253	0.3317	0.1027	0.125	0.7728	0.3096	2.7439
RM	DD	0.2669	0.2719	0.0818	0.125	0.5457	0.3493	2.5272
SEL		0.31030	0.3147	0.0985	0.125	0.8592	0.3663	2.9058

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna). SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

La asimetría es una medida que caracteriza el desvío de la distribución en relación con una total simetría en probabilidad. La existencia de una diferencia importante entre la media y la mediana sugiere asimetría. Dado que en el Cuadro 4 se observa mayoritariamente que la media fue levemente mayor a la mediana, las “colas” de la densidad muestran mayor probabilidad en el extremo derecho (asimetría “sesgada a la derecha”). La diferencia entre la media y la mediana fue mayor para DS (0.0068) y SD (0.0064) en SEL, mientras que para RM fue mayor en DS (0.0054), DD (0.005) y SS (0.0046). Se observó mayor asimetría en SEL (0.0055125) que en RM (0.005), lo que sugiere que la selección aumenta la probabilidad de observar RIBD-AN elevadas. El valor de cría de un animal fundador afecta las decisiones de selección en todas las generaciones subsecuentes (Woolliams et al., 1993); consecuentemente, la selección aumenta la probabilidad de co-selección entre medio-hermanos y hermanos-enteros (Wray y Goddard, 1994, Tabla II, pág. 445) y la consanguinidad. Severson et al. (2019) demostraron recientemente que los apareamientos entre individuos emparentados aumentan la coancestría entre los individuos en la progenie, además de la consanguinidad. La curtosis es una medida de los tamaños combinados de ambos extremos y refleja la probabilidad de las colas. La distribución normal tiene una curtosis de 3. Distribuciones “platicúrticas” tienen una curtosis menor a 3, mientras que distribuciones “leptocúrticas” tienen curtosis mayores a 3. Los resultados en el Cuadro 4 indican que todas las distribuciones de las RIBD-AN fueron “platicúrticas”; es decir, con “colas” más pesadas

que la distribución normal y mayor probabilidad de encontrar “outliers” (datos fuera de rango). Debe señalarse que el algoritmo de estimación fue restringido a valores de $-0.25 \leq \beta_s, \beta_d \leq 0.25$ para evitar que las fracciones de genoma aportadas al individuo por su padre o madre excediesen 0.5 para cada progenitor.

La Figura 3 presenta los gráficos de las estimaciones del “kernel” de la distribución para las RIBD entre SS, DS, SD y DD con el nieto, bajo el modelo de Regresión Ancestral en los datos RM y SEL. En ambos casos la mediana fue mayor que la media: $0.3175 < 0.3233$ para SEL y $0.2676 < 0.2725$ para RM, si bien las diferencias fueron de escasa magnitud relativa.

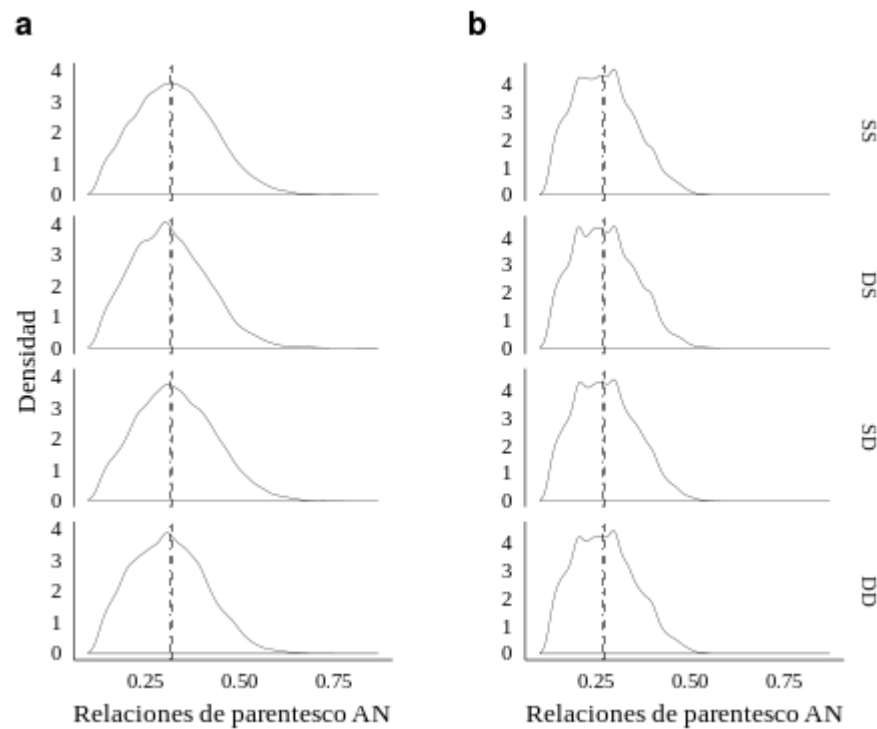


Figura 3. Gráfico de la función de densidad según RIBD-AN: SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna), estimadas bajo el modelo de Regresión Ancestral, sobre: **a.** Datos bajo selección (SEL) y **b.** Datos bajo apareamiento aleatorio (RM). Las líneas verticales para SEL fueron mediana = 0.3175, media = 0.3233; y para RM, mediana = 0.2676 y media = 0.2725.

4.1.4. Distribuciones empíricas por generación

En el Cuadro 5 se presentan la media, el desvío estándar y el coeficiente de variación por generación para las RIBD-AN bajo RM y SEL. Con el fin de evaluar la magnitud del efecto de la selección sobre las RIBD-AN, se estimó la diferencia entre las RIBD-AN promedio en las generaciones 3 y 6 expresándola como porcentaje de la RIBD-AN

promedio en la generación 6, la de mayor valor. Para los datos SEL las proporciones fueron 21.2%, 18.0%, 19.7%, y 15.2%, para SS, DS, SD y DD, respectivamente. Para los datos RM las magnitudes del cambio fueron: SS = 1.6%, DS = 1.8%, SD = 1.7% y DD = 3.9%. Más allá que las diferencias en respuesta al cambio en las RIBD-AN fueron distintas para RM y SEL, no se halló una explicación consistente para el comportamiento de DD en RM, donde el cambio fue sustancialmente mayor que en las otras RIBD-AN. Posteriormente, se analizaron las siguientes diferencias de promedios RIBD-AN entre generaciones subsecuentes: 3 con 4 (3-4), 4 con 5 (4-5) y 5 con 6 (5-6). Para 3-4, 4-5 y 5-6, los cambios observados fueron para RM: SS (-1.6%, 0.2% y 3.0%), DS (-2.3%, 1.2% y 2.9%), SD (-3.9%, 2.9% y 2.6%), y DD (0.0%, -0.1% y 4.1%), mientras que para SEL: SS (3.1%, 10.3% y 9.3%), DS (0.9%, 7.7% y 10.3%), SD (4.0%, 8.5% y 8.6%), y DD (-4.1%, 9.7% y 9.8%). La Figura 4 muestra las funciones de densidad para SS, DS, SD y DD en RM y SEL para cada generación, observándose que los cambios promedios aumentaron en generaciones más recientes. Otro tanto con la dispersión, mostrándose SEL más un poco variable que RM (Cuadro 5 y Figura 4). La interpretación de estos resultados es que el IBD compartido (consanguinidad y coancestría) aumentó por la selección mucho más intensamente en SEL que en RM. En este último caso, el aumento en consanguinidad se debió a deriva génica.

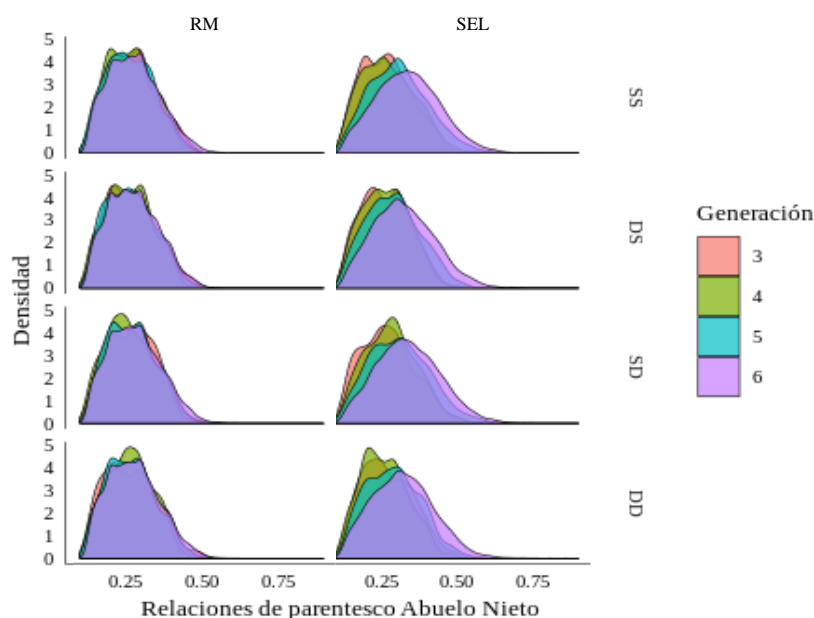


Figura 4. Funciones de densidad por RIBD-AN, donde SS es el abuelo paterno, DS es la abuela paterna, SD es el abuelo materno y DD es la abuela materna, estimadas por el modelo de Regresión Ancestral, sobre datos SEL y RM, para cada una de las generaciones comprendidas en cada réplica.

Cuadro 5. Media, desvío estándar y coeficiente de variación por generación para las distribuciones de las RIBD-AN, estimadas bajo AR sobre RM y SEL.

Data	Tipo	Generación	Media	DS	CV
RM	SS	3	0.2688	0.0787	0.2928
		4	0.2644	0.0752	0.2844
		5	0.2650	0.0776	0.2928
		6	0.2732	0.0815	0.2983
SEL	SS	3	0.2680	0.0809	0.3019
		4	0.2767	0.0879	0.3177
		5	0.3084	0.0961	0.3116
		6	0.3401	0.1040	0.3058
RM	DS	3	0.2690	0.0800	0.2974
		4	0.2628	0.0756	0.2877
		5	0.2659	0.0786	0.2956
		6	0.2738	0.0816	0.2980
SEL	DS	3	0.2665	0.0783	0.2938
		4	0.2689	0.0782	0.2909
		5	0.2913	0.0882	0.3028
		6	0.3249	0.0997	0.3069
RM	SD	3	0.2710	0.0792	0.2923
		4	0.2605	0.0769	0.2952
		5	0.2684	0.0767	0.2858
		6	0.2756	0.0822	0.2983
SEL	SD	3	0.2733	0.0863	0.3158
		4	0.2847	0.0846	0.2972
		5	0.3110	0.0974	0.3132
		6	0.3403	0.1029	0.3024
RM	DD	3	0.2650	0.0844	0.3186
		4	0.2649	0.0742	0.2801
		5	0.2646	0.0783	0.2959
		6	0.2758	0.0825	0.2991
SEL	DD	3	0.2737	0.0826	0.3018
		4	0.2629	0.0763	0.2902
		5	0.2912	0.0881	0.3025
		6	0.3229	0.0997	0.3088

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna). SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

4.1.5. Distribuciones de los estadísticos suficientes

La prueba de normalidad para las distribuciones de los estadísticos suficientes no mostró evidencias sustanciales como para rechazar la hipótesis de normalidad, dado que los p -valores fueron mayores a 0.05. En el Cuadro 6 se presentan el mínimo, el máximo, la mediana, la media, el desvío estándar, la asimetría y la curtosis de las distribuciones de los estadísticos suficientes [$ES1 = \text{Cov}(a_{SS}, a_X) - \text{Cov}(a_{DS}, a_X)$] y $ES2 = [\text{Cov}(a_{SD}, a_X) - \text{Cov}(a_{DD}, a_X)]$ (Cantet et al., 2017), calculadas bajo RM y SEL.

Cuadro 6. Mínimo, máximo, mediana, media, desvío estándar, asimetría y curtosis de las distribuciones de los estadísticos suficientes, estimadas bajo AR sobre RM y SEL.

Datos		Mín.	Máx.	Mediana	Media	DS	Asimetría	Curtosis
RM	ES1	-3.5×10^{-1}	3.6×10^{-1}	8.0×10^{-4}	7.3×10^{-4}	1.3×10^{-1}	2.7×10^{-3}	2.11
	ES2	-3.6×10^{-1}	3.6×10^{-1}	1.4×10^{-3}	2.9×10^{-3}	1.3×10^{-1}	-6.9×10^{-3}	2.13
SEL	ES1	-6.2×10^{-1}	6.4×10^{-1}	1.6×10^{-2}	1.4×10^{-2}	1.7×10^{-1}	3.3×10^{-3}	2.61
	ES2	-6.8×10^{-1}	6.0×10^{-1}	1.7×10^{-2}	1.6×10^{-2}	1.7×10^{-1}	-5.9×10^{-3}	2.59

ES1 = $\text{Cov}(a_{SS}, a_X) - \text{Cov}(a_{DS}, a_X)$. ES2 = $[\text{Cov}(a_{DS}, a_X) - \text{Cov}(a_{DD}, a_X)]$. SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

La distribución de los estadísticos suficientes mostró asimetría en ambos conjuntos de datos. Dado que las medianas fueron menores que las medias (Cuadro 6), el ES1 de RM así como el ES1 y ES2 de SEL mostraron “asimetría a la izquierda”. Asimismo, los estadísticos suficientes empíricos en ambos conjuntos de datos mostraron distribuciones “platicúrtica” (curtosis menores a 3). Finalmente, en función del desvío estándar (DS) de las medias según el sistema de apareamiento, se observó que los datos provenientes de las distribuciones SEL mostraron un mayor grado de dispersión con respecto de sus medias en contraste con las de RM. En resumen, las distribuciones de los estadísticos suficientes bajo SEL fueron asimétrica con mayor probabilidad a la izquierda y más variable que RM.

4.2. Parámetros de la Regresión Ancestral (AR)

4.2.1. Coeficientes de regresión parcial β_s y β_b

La existencia de interacción entre el tipo de apareamiento (RM o SEL) y los bloques (réplicas 3, 4, 5 ó 6) se evaluó mediante la prueba de “aditividad” de Tukey (Rojas, 1973). La hipótesis nula (H_0) fue, que los efectos principales y los bloques se asociaron de manera aditiva dentro del modelo lineal de análisis. La hipótesis alternativa (H_1) consistió en que efectos principales y bloques muestran interacción. La prueba resultó en un p -valor = $0.969 \gg 0.05$, indicando que no hubo evidencia alguna para rechazar H_0 y, por lo tanto, no hubo indicación de la existencia de interacción entre el sistema de apareamiento y réplica. Asimismo, la prueba de normalidad para las distribuciones de β_s y β_b no mostró evidencias suficientes como para rechazar la hipótesis de normalidad al observarse p -valores mayores al tamaño de prueba ($P > 0.05$). Los parámetros de las

distribuciones de β_s y β_D fueron $\beta_s \sim N(0.000821, 0.03644)$ y $\beta_D \sim N(0.000421, 0.02786)$ para los datos RM, y $\beta_s \sim N(-0.00520, 0.02913)$ y $\beta_D \sim N(0.002674, 0.03822)$ para datos bajo SEL.

Cuadro 7. Media y desvío estándar de los coeficientes de regresión parcial: paterno (β_s) y materno (β_D), estimados bajo AR sobre RM y SEL.

Datos	β_s				β_D			
	Media	EE	Asimetría	Curtosis	Media	EE	Asimetría	Curtosis
RM	0.000821	0.000899	0.002465	2.655290	0.000421	0.000786	-0.015406	2.641703
SEL	-0.00520	0.000791	-0.031494	2.51483	0.002674	0.000906	-0.008603	2.519142

SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

En el Cuadro 7 se presentan la media, el error estándar, la asimetría y la curtosis para β_s y β_D según el sistema de apareamiento (RM o SEL). La magnitud de la asimetría para β_s fue mayor en RM, mientras que para β_D la asimetría mayor fue para SEL, lo cual se asocia a una mayor probabilidad en el extremo derecho de la distribución. Por último, los resultados indican que todas las distribuciones de ambos coeficientes de regresión parcial, y para ambos conjuntos de datos fueron “platicúrticas” con curtosis menores a 3, sugiriendo mayor probabilidad en las colas en comparación con las de una distribución normal.

4.2.2. Consanguinidad (F_{AR})

La distribución exponencial fue la que mejor ajustó los tres valores de consanguinidad: calculado con el pedigrí (F_{PED}), con las relaciones realizadas “verdaderas” (F_{GENO}) y bajo Regresión Ancestral (F_{AR}). En el Cuadro 8 se presentan los valores estimados del parámetro λ de la distribución Exponencial para F_{PED} , F_{GENO} y F_{AR} bajo RM y SEL. El Cuadro 9 muestra los valores estimados de $E(X) = 1/\lambda$ y $Var(X) = 1/\lambda^2$ de las distribuciones exponenciales de F_{PED} , F_{GENO} y F_{AR} en datos RM y SEL. Los datos SEL consistieron en 11 684 dúos nieto SS, 11 606 para DS, 11 699 para SD y 11 583 para DD; mientras que para los datos RM se emplearon 11 237 dúos nieto SS, 11 309 para DS, 11 261 para SD y 11 288 para DD, para la estimación de F_{PED} , F_{GENO} y F_{AR} .

Cuadro 8. Parámetro de la distribución exponencial (λ) y su error estándar para los valores de consanguinidad estimadas por pedigrí (F_{PED}), empleando información genómica (F_{GENO}) y por Regresión Ancestral (F_{AR}), según los datos sean RM o SEL.

Data	F_{PED}		F_{GENO}		F_{AR}	
	λ	EE	λ	EE	λ	EE
RM	200.5	0.94	148.5	0.70	152.3	0.71
SEL	23.1	0.11	20.2	0.09	25.1	0.12

SEL (datos provenientes de la distribución con selección), RM (datos provenientes de la distribución con apareamiento aleatorio).

Cuadro 9. Media y Varianza de los valores de F_{PED} , F_{GENO} y F_{AR} para datos RM y SEL, asumiendo distribuciones exponenciales para los parámetros estimados.

Data	F_{PED}		F_{GENO}		F_{AR}	
	E(X)	Var(X)	E(X)	Var(X)	E(X)	Var(X)
RM	5.0×10^{-3}	2.5×10^{-5}	6.7×10^{-3}	4.5×10^{-5}	6.6×10^{-3}	4.3×10^{-5}
SEL	4.3×10^{-2}	1.9×10^{-3}	4.9×10^{-2}	2.5×10^{-3}	4.3×10^{-2}	1.6×10^{-3}

$E(X) = 1/\lambda$; $\text{Var}(X) = 1/\lambda^2$

Esperablemente F_{AR} bajo RM mostró mayores valores de λ y más variables, por otro lado, los mayores valores esperados y varianzas del modelo de probabilidad exponencial se observaron bajo SEL. La interpretación del parámetro de cambio (“rate parameter”) λ es que al aumentar produce caídas más abruptas con menor valor esperado y de mayor variabilidad. Consecuentemente, los valores observados en el Cuadro 8 y 9 son consistentes con el fuerte efecto que tiene la selección sobre la consanguinidad (Woolliams et al., 1993; Wray y Goddard, 1994) al producir mayor probabilidad de encontrar valores de consanguinidad más elevados en datos SEL que bajo RM.

4.2.3. Correlaciones entre las RIBD-AN y los parámetros β_s , β_D y F_{AR}

En los Cuadros 10 y 11 se observan las correlaciones (junto al valor de significancia de la prueba de que el valor estimado es igual a cero), entre las cuatro RIBD-AN, los parámetros β_s , β_D y F_{AR} , para RM y SEL, respectivamente. Las correlaciones entre los cuatro abuelos (SS, DS, SD y DD), entre los abuelos paternos (SS, DS) con β_s y entre los abuelos maternos (SD, DD) con β_D para RM fueron importantes y de magnitudes casi iguales, pero con signo opuesto ($r_{\text{SS}-\beta_s} = 0.961$ y $r_{\text{DS}-\beta_s} = -0.962$; $r_{\text{SD}-\beta_D} = 0.960$ y $r_{\text{DD}-\beta_D} = -0.961$), todas altamente significativas. Estas igualdades en las correlaciones entre ambos

abuelos con el nieto eran esperables dado que las diferencias entre SS y DS y entre SD y DD son los estadísticos suficientes (Cantet et al., 2017) para estimar β_s y β_D , respectivamente. Algo distinto ocurrió con los datos bajo SEL (Cuadro 11) donde las correlaciones fueron $r_{SS-\beta_s} = 0.785$, $r_{DS-\beta_s} = -0.810$, $r_{SD-\beta_D} = 0.787$ y $r_{DD-\beta_D} = -0.826$. Notablemente bajo SEL las correlaciones con las abuelas tuvieron mayor magnitud que con los abuelos, observación escasamente intuitiva dado que la selección ocurrió solamente en los machos. El trabajo de Ma et al. (2015) sugiere una posible explicación a este resultado supuestamente paradójico. Los autores observaron que el número de recombinaciones en machos de la población Holstein en E.E.U.U. mostró una fuerte tendencia a disminuir en el tiempo, asociada con el aumento de la intensidad de selección desde 1960 a nuestros días. Es dable razonar que el aumento en la respuesta a la selección condujo a segmentos del genoma menos variables y, por lo tanto, a una menor probabilidad de detectar recombinaciones. Como resultado de la selección, se hipotetiza que la incapacidad de detectar recombinaciones entre los machos selectos indujo valores de correlaciones con menor magnitud entre las RIBD de los abuelos y los estimadores de β_s y β_D comparado con las RIBD de las abuelas y los estimadores de β_s y β_D .

Cuadro 10. Coeficientes de correlación entre RIBD-AN, coeficientes de regresión parcial paterno (β_s) y materno (β_D) y coeficiente de consanguinidad (F_{AR}) bajo AR sobre RM.

Coeficientes de correlación de Pearson N= 8460. Proba > r suponiendo $H_0 = 0$							
	SS	DS	SD	DD	β_s	β_D	F_{AR}
SS	1.00000	-0.94226	0.02095	0.00635	0.96058	0.00509	0.09008
		<.0001	0.0540	0.5595	<.0001	0.6397	<.0001
DS	-0.94226	1.00000	0.00418	0.01083	-0.96184	-0.00181	0.09845
	<.0001		0.7007	0.3191	<.0001	0.8681	<.0001
SD	0.02095	0.00418	1.00000	-0.93265	0.00430	0.95959	0.10573
	0.0540	0.7007		<.0001	0.6925	<.0001	<.0001
DD	0.00635	0.01083	-0.93265	1.00000	-0.00057	-0.96148	0.08326
	0.5595	0.3191	<.0001		0.9581	<.0001	<.0001
β_s	0.96058	-0.96184	0.00430	-0.00057	1.00000	0.00026	-0.01363
	<.0001	<.0001	0.6925	0.9581		0.9813	0.2101
β_D	0.00509	-0.00181	0.95959	-0.96148	0.00026	1.00000	0.00732
	0.6397	0.8681	<.0001	<.0001	0.9813		0.5009
F_{AR}	0.09008	0.09845	0.10573	0.08326	-0.01363	0.00732	1.00000
	<.0001	<.0001	<.0001	<.0001	0.2101	0.5009	

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna).

Cuadro 11. Coeficientes de correlación entre RIBD-AN, coeficientes de regresión parcial paterno (β_s) y materno (β_D) y coeficiente de consanguinidad (F_{AR}) bajo el AR sobre SEL.

Coeficientes de correlación de Pearson N= 10431. Proba > r suponiendo $H_0 = 0$							
	SS	DS	SD	DD	β_s	β_D	F_{AR}
SS	1.00000	-0.48218 <.0001	0.25367 <.0001	0.17478 <.0001	0.78493 <.0001	0.03247 0.0009	0.43599 <.0001
DS	-0.48218 <.0001	1.00000	0.15959 <.0001	0.07671 <.0001	-0.80919 <.0001	0.02773 0.0046	0.26673 <.0001
SD	0.25367 <.0001	0.15959 <.0001	1.00000	-0.47528 <.0001	0.02969 0.0024	0.78723 <.0001	0.43581 <.0001
DD	0.17478 <.0001	0.07671 <.0001	-0.47528 <.0001	1.00000	0.04086 <.0001	-0.82611 <.0001	0.26900 <.0001
β_s	0.78493 <.0001	-0.80919 <.0001	0.02969 0.0024	0.04086 <.0001	1.00000	-0.00554 0.5716	0.03617 0.0002
β_D	0.03247 0.0009	0.02773 0.0046	0.78723 <.0001	-0.82611 <.0001	-0.00554 0.5716	1.00000	0.04643 <.0001
F_{AR}	0.43599 <.0001	0.26673 <.0001	0.43581 <.0001	0.26900 <.0001	0.03617 0.0002	0.04643 <.0001	1.00000

SS (abuelo paterno), DS (abuela paterna), SD (abuelo materno) y DD (abuela materna).

Capítulo 5: Discusión

Capítulo 5

Discusión

En la presente investigación se evaluaron las posibles distribuciones estadísticas de las relaciones IBD abuelo(a) – nieto(a) y de los parámetros β_s , β_D y F_{AR} de la Regresión Ancestral, para poblaciones bajo apareamiento aleatorio o selección. Dado lo reciente del modelo AR (Cantet et al., 2017) es difícil encontrar otras investigaciones que permitan una comparación directa con los resultados aquí obtenidos. Para evaluar el impacto de la selección en las covarianzas aditivas entre parientes, Tallis y Leppard (1987) encontraron expresiones para las covarianzas entre padre (o madre) e hijo(a), entre hermanos enteros y entre medio hermanos, todas bajo selección. Para ello, emplearon covarianzas condicionales (Tallis y Leppard, 1987; Tallis, 1989) bajo un implícito planteo de normalidad multivariada, dado que sólo con la distribución Gaussiana las covarianzas parciales y condicionales son iguales (Baba et al., 2004). George Michael Tallis refirió a este algoritmo que permite independizar más allá de los VC parentales a los residuos Mendelianos de animales emparentados, como “sistemas de transferencia” (“transfer system” TS, Tallis, 1985). Estos sistemas de transferencia tienen su equivalencia en la literatura de “ecuaciones estructurales” y modelos gráficos (incluyendo “path diagrams”) en el esclarecedor trabajo de Tian (2005), transformando cada ecuación de un sistema de ecuaciones estructurales en una regresión, estrategia que permitió a Cantet et al. (2017) obtener el modelo AR. Más aún, la expresión “regresión ancestral” aparece ya en Tallis (1987) y en Tallis (1989, sección 3, página 166). La inferencia sobre las distribuciones de los parámetros del modelo de la Regresión Ancestral genera conocimiento para la estimación paramétrica, primer paso para ajustar AR en las evaluaciones genéticas.

5.1. Relaciones de IBD compartido abuelo-nieto

El primer aporte importante de esta tesis es el haber encontrado distribuciones empíricas de las RIBD-AN bajo el modelo AR. Se observó que la distribución Beta fue la que mejor ajustó los datos simulados, utilizando como criterio el máximo del logaritmo de la función de verosimilitud (Cuadro 1 del Capítulo 4). Para modelar la distribución de una RIBD- AN, la Beta tiene la ventaja de que su espacio muestral está limitado al

intervalo $[0,1]$ al igual que la RIBD-AN en esta investigación. Comparativamente, la distribución normal no tiene limitación en su espacio que es toda la recta real $(-\infty, \infty)$; asimismo, las distribuciones Log-normal, Weibull, Gamma y Exponencial no tienen límite por derecha encontrándose sus espacios muestrales definidos en el intervalo $(0, \infty)$. Walters y Cannings (2005) calcularon la distribución de la longitud del IBD compartido entre abuelo y nieto, bisabuelo y bisnieto, y tatarabuelo-tataranieta. Su figura 2 (página 60) sugiere que la distribución Beta modelará razonablemente bien las distintas covarianzas ancestrales obtenidas por los citados autores analíticamente.

Las distribuciones de RIBD-AN bajo SEL tuvieron menor magnitud en los valores estimados de α y β (SS = 6.53, 13.22, DS = 6.84, 14.77, SD = 6.72, 13.53, DD = 6.77, 14.74) que sus contrapartes bajo RM (SS = 8.12, 21.75, DS = 8.09, 21.61, SD = 8.05, 21.33, DD = 7.90, 21.15) (Cuadro 2, Capítulo 4), además, los estimadores del valor esperado y la varianza de las RIBD-AN para SEL fueron en promedio 1.2 veces superiores a los de RM (Cuadro 3, Capítulo 4). Nótese que el parámetro α fue siempre de menor magnitud que β (Cuadro 2, Capítulo 4). En términos de error estándar del estimador, las medias de α y β no discreparon respecto al valor que se está estimando, dado que rara vez superaron 0.20 para SEL, mientras que los errores estándar de los estimadores con RM mostraron una mayor distancia entre la media poblacional ($\alpha/(\alpha + \beta)$) y la muestral. Una característica saliente de la distribución Beta, es la de permitir ajustar densidades con colas desiguales (“skewness”) sobre la base de los valores estimados de α y β (Nadarajah y Gupta, 2004). En el Cuadro 2 del capítulo 4 se observa que las distribuciones fueron *asimétricas hacia la derecha*, porque hubo más valores de RIBD-AN superiores a 0.35 que inferiores a 0.15 (promedio del modelo clásico sin consanguinidad igual a 0.25). En cuanto a la dinámica del efecto de la selección sobre las RIBD-AN en AR, se calcularon las diferencias entre las medias de distintas generaciones de selección expresándolas en porcentaje sobre el valor anterior (Cuadro 5, del capítulo 4). Recordando que las generaciones 1 y 2 no tuvieron información genómica como para calcular las RIBD-AN, se obtuvo para SEL un valor mayor a 9 % de diferencia a partir de la generación 4 (segundo cambio generacional), siendo 18.5% la magnitud relativa del cambio entre las generaciones 3 y 6. De manera opuesta, los datos RM mostraron diferencias entre generaciones mucho menores en el rango de -2.02% a 3.14% , siendo de 2.24% entre las generaciones 3 y 6.

En resumen, las RIBD-AN empíricamente obtenidas siguieron distribuciones Betas con parámetros $\alpha \gg 1$, $\beta \gg 1$ y $\alpha \ll \beta$ y con una importante asimetría hacia la derecha (Figura 3, del capítulo 4). Asimismo, las distribuciones de los datos SEL mostraron mayor valor esperado y varianza que los datos RM. Estos resultados son trascendentes al momento de estimar los parámetros del modelo AR, particularmente cuando las abuelas no están genotipadas. Dado el costo del panel denso en SNP por animal, los criadores están menos dispuestos a genotipar hembras, las cuales son más frecuentes en las poblaciones núcleos de selección. En tales casos, las RIBD abuela-nieto podrán estimarse por muestreo de la distribución de las diferencias de dos Betas (Pham-Gia y Turkkan, 1993) o, mejor aún, por normalidad, empleando los valores de las covarianzas entre el individuo cuyos parámetros deben ser estimados, con sus tíos (hijos de la abuela) o tíos abuelos (medio hermanos paternos de la abuela), como datos para contrastar en el cálculo del estadístico suficiente (RIBD abuelo nieto menos RIBD abuela nieto).

5.2. Parámetros del modelo de Regresión Ancestral (AR)

Otro hallazgo importante de esta tesis fue el haber observado distribuciones normales para los parámetros β_s y β_D de la Regresión Ancestral. Los estimadores del kernel para los datos RM fueron $\beta_s \sim N(0.000821, 0.03644)$ y $\beta_D \sim N(0.000421, 0.02786)$, mientras que las distribuciones estimadas bajo SEL produjeron $\beta_s \sim N(-0.00520, 0.02913)$ y $\beta_D \sim N(0.002674, 0.03822)$. Los valores distintos de cero para las cuatro medias estimadas reflejan el hecho que la probabilidad de transmitir el genoma sin recombinación alguna en hombres es menor a 10^{-21} y en mujeres menor a 10^{-28} (Graham Coop, 2013, <https://gcbias.org/2013/10/20/how-much-of-your-genome-do-you-inherit-from-a-particular-grandparent/>). Dado que β_s y β_D , reflejan las recombinaciones en los genomas de las parejas de abuelos que se hallan presentes en las gametas parentales y que constituyen al individuo, es muy factible que las medias sean distintas de cero. Los valores absolutos de β_s y β_D fueron significativamente mayores bajo SEL que con RM (Cuadro 7, del capítulo 4). Como Albrechtsen et al. (2015) probaron analíticamente que en una población exogámica (que no puede autofertilizarse) la selección aumenta la cantidad de IBD compartido, la mayor magnitud de β_s y β_D se puede asociar a mayores valores de RIBD abuelo-nieto en presencia de selección, dado que el mínimo valor de la RIBD abuelo -nieto es igual a 0.25 más o menos β_s (o β_D). Asimismo, la selección aumentó la variabilidad en las estimaciones de β_D parámetros.

Por último, λ para el coeficiente de consanguinidad bajo AR (F_{AR}) fue siempre mayor y su dispersión fue mucho más elevada bajo RM que bajo SEL (Cuadro 8, Capítulo 4). Nótese que F_{AR} (ecuación [2.4], Capítulo 2), contempla las diferencias entre las RIBD de los abuelos paternos y maternos con la madre y el padre del individuo, multiplicados por los respectivos coeficientes de regresión parcial; además, incluye las RIBD “cruzadas” entre los abuelos paternos y maternos por los coeficientes de regresión parcial de los padres. Esta expresión generaliza la fórmula clásica para el cálculo de la consanguinidad por pedigrí y le confiere un comportamiento intermedio en el cálculo entre los valores de F por pedigrí y aquellos calculadas utilizando las relaciones realizadas “verdaderas” o en el caso de las “corridas de homocigosidad” (ROH) en la secuencia, particularmente, bajo procesos de selección (Forutan et al., 2018). Esto implica, que las estimaciones de consanguinidad mediante F_{AR} se ubicarán entre las obtenidas por pedigrí (F_{PED}) y aquellas más extremas obtenidas mediante ROH que tienden a sobrestimar, debido a que estos últimos métodos detectan “falsos positivos” (exceso de segmentos homocigotas que ocurren al azar, Forutan et al, 2018). Volviendo al problema de detección de las recombinaciones en poblaciones bajo selección discutido previamente, Forutan et al. (2018) observaron que la sobrestimación de F fue menor en las generaciones más antiguas debido a que la ocurrencia al azar de pequeños segmentos ROH depende del número de eventos de recombinación. Nótese que ROH es un método no-paramétrico de detección de F , mientras que AR es un método paramétrico. Consecuentemente, la menor capacidad de detectar recombinación en poblaciones selectas al emplear información genómica tiende a subestimar el valor de RIBD-AN cuando los machos son selectos y no las hembras y a sobrestimar la consanguinidad IBD, a menos que se utilice un método paramétrico que combine la información del pedigrí y la genómica para calcular IBD, como es en el caso de AR. En general, la selección aumentó la magnitud y variabilidad del material genómico compartido IBD, tanto en las relaciones de parentesco como en la consanguinidad.

Capítulo 6: Conclusiones

Capítulo 6

Conclusiones

La originalidad de esta tesis consiste en haber encontrado mediante simulación estocástica las distribuciones de probabilidad para los parámetros de la Regresión Ancestral (β_s , β_D), así como para las relaciones de RIBD-AN, los estadísticos suficientes para estimar β_s y β_D y para el coeficiente de consanguinidad F_{AR} . Las principales conclusiones son las siguientes:

1. Se evaluaron distintas distribuciones continuas para modelar las relaciones de identidad por descendencia bajo la Regresión Ancestral entre abuelos y nieto(a) (RIBD-AN) y se observó que la Beta presentó el mejor ajuste al utilizar como criterio el máximo del logaritmo de la función de verosimilitud.
2. Se encontró que los coeficientes de regresión parcial paterno (β_s) y materno (β_D), así como también los estadísticos suficientes (diferencia entre las RIBD abuelo – nieto y abuela – nieto) siguieron una distribución normal.
3. Cuando se practicó selección (SEL) las distribuciones RIBD-AN mostraron asimetría hacia la derecha, acompañadas de una mayor magnitud en la media y dispersión, al compararlas con sus contrapartes bajo apareamiento al azar (RM).
4. La distribución que mejor ajustó los valores estimados de consanguinidad fue la Exponencial. Las estimaciones del parámetro de cambio λ fueron menores y menos variables con SEL que con RM, con lo cual valores esperados y varianzas bajo SEL fueron mayores que bajo RM. Las estimaciones de consanguinidad por Regresión Ancestral (F_{AR}), mostraron valores intermedios entre aquellas obtenidas por pedigrí (F_{PED}) y las estimaciones F_{GENO} empleando las relaciones realiza “verdaderas”. En general, la selección aumentó la magnitud y variabilidad del material genómico compartido IBD, tanto en las relaciones de parentesco como en la consanguinidad.

Las distribuciones aquí obtenidas servirán de base para estimar mediante máxima verosimilitud o métodos Bayesianos, los parámetros de la Regresión Ancestral: un modelo genético cuantitativo paramétrico y causal, que permite considerar el efecto de la recombinación y la herencia en segmentos. Siendo además Markoviano, permite la inversión simple de Σ tal como C. R. Henderson logró con la matriz de relaciones A .

Todas estas propiedades hacen de AR una sólida herramienta teórica para la predicción del valor de cría de cientos de millones de animales de manera computacional altamente competitiva.

BIBLIOGRAFÍA

- Abney, M., McPeck, M. S. y Ober, C. 2000. Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.* 66: 629–650.
- Albrechtsen, A., Moltke, I. y R. Nielsen. 2010. Natural selection and the distribution of identity-by-descent in the human genome. *Genetics.* 186: 295–308.
- Baba, K., Shibata, R. y Sibuya, M. 2004. Partial correlation and conditional correlation as measures of conditional independence. *Aust N Z J Stat.* 46: 657 - 664.
- Berman, A. y Plemmons, R. J. 1994. *Non-negative matrices in the Mathematical Sciences.* SIAM, Philadelphia, USA.
- Bickeböllner, H. y Thompson, E. A. 1996. Distribution of genome shared IBD by half-sibs: approximation by the Poisson clumping heuristic. *Theor. Populat. Biol.* 50: 66 - 90.
- Bulmer, M. G. 1985. *The mathematical theory of quantitative genetics.* Oxford: Clarendon Press, NY, USA.
- Buonaccorsi, J. P. 2010. *Measurement Error: Models, Methods, and Applications.* Chapman & Hall, London, UK.
- Cantet, R. J. C., García-Baccino, C. A., Rogberg-Muñoz, A., Forneris, N. S. y Munilla, S. 2017. Beyond genomic selection: The animal model strikes back (one generation). *J. Anim.Breed. Genet.* 134: 224 - 231.
- Cantet, R. J. C. y Vitezica, Z. G. 2014. Properties of Mendelian residuals when regressing breeding values using a genomic covariance matrix. 10th WCGALP. Vancouver (Canada), 17–22 August 2014. Paper number 687.
- Chen, S. X. 1999. A beta kernel estimation for density functions. *Computational Statistics and Data Analysis.* 31: 131 - 145.
- Cockerham, C. C. y Weir, B. S. 1969. Pedigree mating with two linked loci. *Genetics.* 61: 923-940.
- De los Campos, G., Sorensen, D. y Gianola, D. 2015. Genomic Heritability: What Is It? *PLOS Genetics.* DOI:10.1371/journal.pgen.1005048.
- Donnelly, K. 1983. The probability that related individuals share some section of genome identical by descent. *Theoret. Pop. Biol.* 23: 34 - 64.
- Emik, L. O. y Terrill, C. R. 1949. Systematic procedures for calculating inbreeding coefficients. *J. Hered.* 40: 51 - 55.

- Falconer, D. S. 1960. *Genética cuantitativa*. Compañía Editorial Continental, D. F., México.
- Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinburgh*. 52: 399-433.
- Forneris, N. S., Legarra, A., Vitezica, Z. G., Tsuruta, S., Aguilar, I., Misztal, I. y Cantet, R.J.C. 2015. Quality control of genotypes using heritability estimates of gene content at the marker. *Genetics*. 199: 675 - 681.
- Forneris, N. S. 2016. Relaciones genómicas y control de calidad del genotipado empleando la noción de identidad por descendencia en el marcador (tesis doctoral). Universidad de Buenos Aires, C.A.B.A, Argentina.
- Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F. S. y Sargolzaei, M. 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American holstein cattle. *BMC Genomics*. 19: 98.
- Fuller, W. A. 1987. *Measurement error models*. Wiley, NY, USA.
- Gianola, D., De los Campos, G., Hill, W. G., Manfredi, E. y Fernando, R. L. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics*. 183: 347-363.
- Guo, S.W. 1995. Proportion of genome shared identical by descent by relatives: concept, computation, and applications. *American Journal of Human Genetics*. 56: 1468 -1476.
- Han, L. y Abney, M. 2011. Identity by descent estimation with dense genome-wide genotype data. *Genet. Epidemiol.* 35: 557 - 567.
- Harville, D. A. 2008. *Matrix Algebra from a Statistician's Perspective*. Springer, NY, USA.
- Hazel, L. N. 1943. The genetic basis for constructing selection indexes. *Genetics*. 28: 476-490.
- Henderson, C. R. 1963. *Selection index and expected genetic advance*. NAS-NRC, Natl. Acad. Sci., WA DC, USA.
- Henderson, C. R. 1973. Sire evaluation and genetic trends. *Journal of Animal Science*. (Symposium). 10-41.
- Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 31: 423-447.
- Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationships matrix used in prediction of breeding values. *Biometrics*. 32: 69 - 83.
- Henderson, C. R. y Quaas, R. L. 1976. Multiple trait evaluation using relative records. *J. Anim. Sci.* 43: 1188 - 1197.

- Henderson, C. R. 1984. *Applications of linear models in animal breeding*. University of Guelph, Guelph, Ontario, Canada.
- Im, S., Fernando, R. L. y Gianola, D. 1989. Likelihood inferences in animal breeding under selection: a missing-data theory view point. *Genet. Sel. Evol.* 21: 399 - 414.
- Jones, M. C., Marron, J. S. y Sheather, S. J. 1996. A Brief Survey of Bandwidth Selection for Density Estimation. *J. Am. Stat. Assoc.* 91: 401 - 407.
- Jones, M. C. 1990. Variable kernel density estimated and variable kernel density estimates. *Austral. J. Statist.* 32: 361 - 371.
- Kenward, M. G. y Roger J. H. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics.* 53: 983 - 997.
- Lange, K. 1978. Central limit theorems of pedigrees. *J. Math. Biol.* 6: 59-66.
- Legarra, A., Aguilar, I. y Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92: 4656-4663.
- Lush, J. L. 1935. Progeny test and individual performance as indicators of an animal's breeding value. *J. Dairy Sci.* 18: 1.
- Lush, J. L. 1945. *Animal Breeding Plans*. 402 pp. The Collegiate Press Inc. Ames, IA, USA.
- Lynch, M., y Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA, USA.
- Ma, L., O'Connell, J. R., VanRaden, P. M., Shen, B., Padhi, A., Padhi, A., Sun, Ch., Bickhart, D., Cole, J., Null, D. J., Liu, G., Da, Y., y Wiggans, G. R. 2015 Cattle sex-specific recombination and genetic control from a large pedigree analysis. *PLoS Genet.* 11: e1005387.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Masson & Co, Paris, Francia.
- Malécot, G. 1969. *The mathematics of heredity*. W. H. Freeman, San Francisco, USA.
- Meuwissen, T. H., Hayes, B. J. y Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 157: 1819 - 1829.
- Molinuevo, H. A., y J. L. Lush. 1964. Reliability of first, second, and third records for estimating the breeding value of dairy cows. *J. Dairy Sci.* 47: 890.
- Munilla, S. y Cantet, R. J. C. 2012. Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters - an application to the maternal animal model. *J. Anim. Breed. Genet.* 129: 173 - 187.

- Nadarajah, S. y Gupta, A. K. 2004: Characterizations of the Beta Distribution. *Communications in Statistics - Theory and Methods*. 33: 2941 - 2957.
- Pearl, J. 2000. *Causality: models, reasoning, and inference*. Cambridge University Press, NY, USA.
- Pham-Gia, T. y Turkkan, N. 1993. Bayesian analysis of the difference of two proportions. *Commun. Statist. Theor. Meth.* 22: 1755 - 1771.
- Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. *Biometrics*. 32: 949 - 953.
- Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71: 1338 - 1345.
- Rojas, B. A. 1973. On tukey's test of additivity. *Biometrics*. 29: 45 - 52.
- Roverato, A. 2000. Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika*. 87: 99 - 112.
- Sargolzaei, M. y Schenkel, F. S. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics*. 25: 680 - 681.
- Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123: 218-223.
- Schnell, F. W. 1961. Some general formulations of linkage effects in inbreeding. *Genetics*. 46: 947-957.
- Severson, A. L., Carmi, S. y Rosenberg, N. A. 2019. The effect of consanguinity on between-individual identity-by-descent sharing. *Genetics*. 212: 305 - 316.
- Silverman, B. W. 1998. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC, NY, USA.
- Tallis, G. M. 1985. Transfer systems and covariance under assortative mating. *Theor. Appl. Genet.* 70: 497 - 504.
- Tallis, G. M. 1987. Ancestral covariance and the Bulmer effect. *Theor. Appl. Genet.* 73: 815 - 820.
- Tallis, G. M. y Leppard, P. 1987. The joint effects of selection and assortative mating on a single polygenic character. *Theor. Appl. Genet.* 75: 41 - 45.
- Tallis, G. M. 1989. The effects of selection and assortative mating on genetic parameters. *J. Anim. Breed. Genet.* 106: 163 -179.
- Tian, J. 2005. Identifying direct causal effects in linear models. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI)* 346-353. AAAI press, Palo Alto, CA, USA.

- Walters, K. y Cannings, C. 2005. The probability density of the total IBD length over a single autosome in unilineal relationships. *Theor. Popul. Biol.* 68: 55 - 63.
- Woolliams, J. A., Wray, N. R. y Thompson, R. 1993. Prediction of long-term contributions and inbreeding in populations undergoing mass selection. *Genet. Res.* 62: 231-242.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* 56: 330 - 338.
- Wray, N. R. y Goddard, M. E. 1994. Increasing long-term response to selection. *Genet. Sel. Evol.* 26: 431-451.