



SIMONE ROUSSELET LAFRATTA

TRANSMISSÃO DE HERANÇA GENÉTICA: UM MODELO MATEMÁTICO

Santo André, 2019



UNIVERSIDADE FEDERAL DO ABC

CENTRO DE MATEMÁTICA, COMPUTAÇÃO E COGNIÇÃO

SIMONE ROUSSELET LAFRATTA

TRANSMISSÃO DE HERANÇA GENÉTICA: UM MODELO MATEMÁTICO

Orientador: Prof. Dr. Prof. Rafael de Mattos Grisi

Dissertação de mestrado apresentada ao Centro de
Matemática, Computação e Cognição para
obtenção do título de Mestre

ESTE EXEMPLAR CORRESPONDE A VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELA ALUNA SIMONE ROUSSELET LAFRATTA,
E ORIENTADA PELO PROF. DR. PROF. RAFAEL DE MATTOS GRISI.

SANTO ANDRÉ, 2019

Sistema de Bibliotecas da Universidade Federal do ABC
Elaborada pelo Sistema de Geração de Ficha Catalográfica da UFABC
com os dados fornecidos pelo(a) autor(a).

Rousselet Lafratta, Simone
Transmissão de Herança Genética : Um Modelo Matemático / Simone
Rousselet Lafratta. — 2019.

76 fls. : il.

Orientador: Rafael de Mattos Grisi

Trabalho de Conclusão de Curso — Universidade Federal do ABC,
Mestrado Profissional em Matemática em Rede Nacional - PROFMAT,
Santo André, 2019.

1. probabilidade. 2. genética. 3. recorrência. I. de Mattos Grisi, Rafael.
II. Mestrado Profissional em Matemática em Rede Nacional -
PROFMAT, 2019. III. Título.



MINISTÉRIO DA EDUCAÇÃO
Fundação Universidade Federal do ABC

Avenida dos Estados, 5001 – Bairro Santa Terezinha – Santo André – SP
CEP 09210-580 · Fone: (11) 4996-0017

FOLHA DE ASSINATURAS

Assinaturas dos membros da Banca Examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado da candidata, SIMONE ROUSSELET LAFRATTA realizada em 11 de Novembro de 2019:



Prof.(a) GLEICIANE DA SILVA ARAGÃO
UNIVERSIDADE FEDERAL DE SÃO PAULO



Prof.(a) SINUE DAYAN BARBERO LODOVICI
UNIVERSIDADE FEDERAL DO ABC

Prof.(a) EDUARDO GUERON
UNIVERSIDADE FEDERAL DO ABC

Prof.(a) RENATO ALESSANDRO MARTINS
UNIVERSIDADE FEDERAL DE SÃO PAULO



Prof.(a) RAFAEL DE MATTOS GRISI
UNIVERSIDADE FEDERAL DO ABC - Presidente

* Por ausência do membro titular, foi substituído pelo membro suplente descrito acima: nome completo, instituição e assinatura

Este exemplar foi revisado e alterado em relação à versão original, de acordo com as observações levantadas pela banca no dia da defesa, sob responsabilidade única do(a) autor(a) e com a anuência do(a) orientador(a).

Santo André/SP



05 de Fevereiro de 2020

Assinatura do(a) autor(a):

Simone Zauschet Sakatte

Assinatura do(a) orientador(a):

Rafael

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoas de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

AGRADECIMENTOS

Agradeço aos meus pais que sempre me ofereceram o melhor estudo durante toda minha vida, ao meu marido que nunca deixou eu desistir , ao meu orientador que me ensinou o que é ser um professor de verdade e claro a Deus que permitiu que eu chegasse até aqui com saúde.

RESUMO

O estudo da Genética de Populações depende fortemente da Matemática, utilizando noções de Probabilidade, Estatística e Equações de Recorrências como ferramentas principais na modelagem. Neste trabalho introduzimos algumas noções de probabilidade e equações de recorrências, e analisamos um modelo básico de transmissão de hereditariedade para uma população dióica.

Este é um estudo do artigo: “ Modelos Matemáticos na genética de Populações”, publicado na Revista Eletrônica Paulista e Matemática em Dezembro de 2018 por Vinícius Freitas de Oliveira e Suzete Maria Silva Afonso, ambos da UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Palavras-chave: probabilidade, recorrência, equilíbrio, alelo, genética

ABSTRACT

The study of Population Genetics depends heavily on Mathematics, using notions of Probability, Statistics and Recurrence Equations as main tools in modeling. In this work we introduce some notions of probability and recurrence equations, using those to analyze a basic model of transmission of heredity to a dioecious population.

This is a study of the article: “Mathematical Models in Population Genetics”, published in Revista Eletrônica Paulista e Matemática in December 2018 by Vinícius Freitas de Oliveira and Suzete Maria Silva Afonso, both from UNESP - Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Keywords: probability, recurrence, equilibrium, allele, genetics

CONTEÚDO

Introdução	1
1 BREVE INTRODUÇÃO DA GENÉTICA	3
1.1 A importância da Genética	5
1.1.1 A genética na medicina	5
1.1.2 A genética na agricultura	6
1.2 Noções básicas para introdução à Genética	6
1.3 Genética e Sexo	10
1.3.1 Cromossomos sexuais e determinação genética do sexo	11
1.3.2 Herança ligada aos cromossomos sexuais	13
2 FERRAMENTAS MATEMÁTICAS	19
2.1 Equações de Diferença	20
2.1.1 Sequências definidas por recorrência	21
2.1.2 Recorrências Lineares de Primeira Ordem	23
2.1.3 Equações Lineares de Segunda Ordem	29
2.2 Probabilidade	36
2.2.1 Teoria da probabilidade	37
2.2.2 Definição de Probabilidade	38
2.2.3 Propriedades de uma Probabilidade	41
2.2.4 Probabilidade Condicional	41
3 MODELOS PARA TRANSMISSÃO DE HERANÇA GENÉTICA	45
3.1 Frequência genotípicas e alélicas	45
3.2 Hipóteses Iniciais do Modelos	48
3.2.1 População Monóica ou Dióica	48
3.2.2 Cromossomos Autossômicos ou Sexuais	49
3.2.3 Acasalamento Aleatório	49
3.2.4 População Infinita	53
3.2.5 Gerações não-sobrepostas	54

3.2.6 Ausência ou Presença de Mutação, Seleção ou Migração	54
3.3 Modelagem	55
3.3.1 Modelo básico	55
3.4 Modelo Básico para População Díóica	59
3.5 Modelo para genes associados a cromossomos sexuais	62
Bibliografia	65

INTRODUÇÃO

O estudo da genética e hereditariedade faz parte do passado do presente (e claramente do futuro) de diversas áreas do conhecimento. Cada vez mais cientistas dedicam seus projetos em prol de novas descobertas que possam evoluir esta ciência para o avanço da Medicina, da Agricultura, da Veterinária, da Pecuária, e em muitas outras áreas. Desde o mapeamento do genoma humano e das controversas pesquisas com células tronco, até o melhoramento genético de plantas, criando espécies mais resistentes e duradouras e frutas, legumes e verduras, os avanços no campo da genética marcaram de maneira definitiva a ciência do último século.

São muitas ciências se empenham neste avanço, e dentre elas a Matemática possui um papel importante. Iniciado por Mendel no século 19, os modelos matemáticos e estatísticos possuem um papel importante no entendimento herança genética.

E tais ideias estão presentes também na sala de aula. Os primeiros conceitos de Genética são introduzidos no 3º Ano do Ensino Médio, onde é apresentado todo o básico sobre cromossomos e suas divisões celulares até atingir o crossing over, onde finalmente a matemática faz sua primeira aparição no estudo de Biologia, com os cálculos de probabilidade de transmissão de um alelo. Vale ressaltar que no 2º Ano do Ensino Médio já foi ensinado combinatória e probabilidade básicas, de modo que os alunos percebam a interdisciplinaridade entre as matérias.

Mas é possível fazer mais com conceitos não mais complicados que PA's e PG's, também estudadas no Ensino Médio. Além da probabilidade envolvida na genética básica, existem Modelos Matemáticos simples que permitem que entendamos como essas ideias básicas de hereditariedade geram o equilíbrio de proporções genotípicas em uma população.

Uma reclamação constante sobre o conteúdo de matemática na escola, principalmente no Ensino Médio, é sobre a desconexão deste com as demais áreas. Diversos conceitos apresentados neste período, como matrizes e progressões, parecem não ter utilidade fora de

si mesmo. Os problemas aplicados são geralmente simples e irreais. Não parecem guardar conexão com nenhuma outra área ou conteúdo estudado nos demais anos.

Mesmo em ciências aplicadas de viés mais matemático, como a Física, esta conexão é pouco explorada. O aluno recebe a informação de como um fenômeno se comporta, faz alguns experimentos em laboratório para verificar a validade, mas raramente discute o passo inicial de todo este processo: a modelagem.

Mostrar para os alunos o papel da Matemática na modelagem de processos naturais é fundamental para a compreensão do método científico. Introduzir tais conceitos no estudo da Genética é uma ferramenta para tornar o processo modelagem acessível a ao menos parte dos alunos, em particular àqueles se interessam mais na área tecnológica e aplicada do que na teórica.

A intenção deste trabalho é o estudo de modelos que utilizem ferramentas simples, mas que expressem conceitos importantes. Conceitos estes que futuramente poderão ser adaptados para atividades em sala de aula, que ajudariam na compreensão de conceitos matemáticos, biológicos e principalmente do processo de “tradução” de um problema natural (no caso biológico) em um problema matemático. Por isso a importância da ênfase nas hipóteses de cada modelo, e como cada uma delas influencia no modelo.

No primeiro capítulo apresentaremos conceitos básicos da Genética necessários para os modelos estudados posteriormente. Discutiremos algumas definições básicas, incluindo exemplos de algumas doenças hereditárias.

No segundo capítulo faremos a apresentação de ferramentas matemáticas necessárias para o estudo dos modelos propostos, tais como recorrências e probabilidade.

Para finalizar, no terceiro capítulo, apresentaremos modelos básicos de transmissão de herança genética para espécies monóicas e dióicas. Apresentaremos também a noção de equilíbrio de Hardy-Weinberg, e como os modelos apresentados levam naturalmente à tal estado.

BREVE INTRODUÇÃO DA GENÉTICA

Genética é o nome dado ao estudo da hereditariedade, o processo pelo qual as características são passadas dos genitores para a prole de modo que todos os descendentes se assemelham a seus ancestrais. Genética pode ser também conceituada como o estudo dos genes.

Apesar de alguns a definirem hoje como o estudo da hereditariedade, esse fenômeno já despertava o interesse do homem muito antes da biologia ou da genética existirem como disciplinas científicas. No passado, a melhoria de plantas e animais domésticos era feita selecionando indivíduos desejáveis para se cruzar. Os povos antigos também deveriam ter se perguntado: Por que as crianças se parecem com seus pais? No entanto estas pessoas não podiam ser chamadas de geneticistas.

A palavra genética vem de gene, e os genes são o foco do assunto, não importando se o estudo é feito em nível molecular, celular, morfológico, populacional ou evolutivo, os genes são sempre centrais em qualquer estudo.

A genética envolve o estudo de dois fenômenos distintos e complementares: a hereditariedade e a variação. O primeiro está relacionada a tendência de igual gerar igual, ou seja, explica o fato pelo qual os descendentes (filhos) se assemelham aos seus ascendentes (pais e avós). Em contrapartida, a variação pode ser definida como todas as diferenças genéticas ou ambientais entre organismos relacionados pela descendência. Tais variações podem ser, portanto, devidas ao meio, não-hereditárias, como também podem ser produzidas por alterações na constituição genética, neste caso, hereditárias.

A hereditariedade e a variação são forças antagônicas, pois enquanto a hereditariedade está relacionada com a semelhança entre os indivíduos no decorrer das gerações, a variação é exatamente ao contrário.

Provavelmente a genética tenha despertado a atenção do homem há muitos anos. Existem evidências que há mais de 10.000 anos, o homem já se preocupava em selecionar plantas e animais para a sua sobrevivência.

Muitas hipóteses foram formuladas para explicar a transmissão das características hereditárias ao longo do tempo. O primeiro estudo realizado das características herdadas foi na Áustria pelo monge Gregor Mendel entre 1860-1870. Ele concluiu que as células reprodutivas são formadas sem qualquer plano pré-definido. Em vez disso, um determinante do pai e outro da mãe com diferentes características é transmitido de forma independente.

Infelizmente o trabalho de Mendel só foi reconhecido em 1900, 16 anos após sua morte (1884). Assim, 1900 foi considerado o marco zero, ou o ano do nascimento da genética. É por essa razão que ela é conhecida como uma ciência do século XX.

A genética é, portanto, uma ciência relativamente nova, mas que tem evoluído espetacularmente, sobretudo porque despertou a atenção de vários ramos do conhecimento humano.

Nos primeiros anos, o estudo da genética estava voltado para o controle genético dos caracteres, mas a partir dos anos 50 foram intensificadas as pesquisas sobre a natureza química do gene, seu funcionamento e regulação que contribuíram para o desenvolvimento de uma nova área da tecnologia – a Engenharia Genética.

Na área da Matemática, os primeiros estudos voltados para os comportamentos genéticos foram realizados por Francis Galton (1822-1911) que realizou aplicações da Estatística na Biologia, chamado Genética das Populações. É possível, assim, através dela, organizar programas para a conservação de espécies ameaçadas, desenvolver levantamentos populacionais de portadores de genes com doenças e identificar genes propícios à doenças em humanos.



Figura 1: DNA organizado em dupla hélice

1.1 A IMPORTÂNCIA DA GENÉTICA

A genética é uma das principais ciências a serviço do homem. Ela tem contribuído em várias áreas de atividades, como na biologia, medicina, agronomia, veterinária, entre outras, muitas aplicações com fins mais científicos e outros com finalidades comerciais. A seguir alguns exemplos da importância da genética.

1.1.1 *A genética na medicina*

A genética teve e tem um profundo impacto na medicina. Embora as conexões entre algumas doenças e a herança tivessem sido feitas há séculos, as mais importantes ligações entre os novos princípios mendelianos descobertos e as doenças foi publicado em 1902.

Os pesquisadores fizeram grandes esforços para estabelecer ligações entre genes defeituosos e doenças. Os genes nas pessoas normais são chamados de genes tipo selvagem; aqueles com defeitos são chamados de genes mutantes. Eles resultam em condições anormais ou mutantes. Os defeitos em genes mutantes são chamados de mutações.

A hemofilia, ou doença do sangramento, foi um dos primeiros distúrbios humanos a serem ligados a um gene mutante e ao seu produto funcional. No passado, os indivíduos com esse defeito herdado, de coagulação do sangue, morriam durante a infância. Hoje em dia, eles vivem uma vida normal graças a tratamentos eficazes

com fator de coagulação produzido por culturas de células de mamíferos geneticamente modificadas.

1.1.2 *A genética na agricultura*

Além de seu impacto na medicina, a genética teve um tremendo impacto na agricultura. Uma das grandes conquistas foi a aplicação dos princípios mendelianos ao desenvolvimento do milho híbrido .

No caso do Brasil, a genética de plantas e animais também assume um papel muito importante, porque há uma forte pressão no incremento na produção de alimentos, tendo como objetivos melhorar o padrão de nutrição do povo brasileiro. Empresas Nacionais e Universidades Públicas desenvolvem programas de melhoramento genético para diferentes espécies (milho, feijão, soja, algodão, café, amendoim) contribuindo para o aumento da produção, melhoria da qualidade nutricional, resistência a pragas e doenças entre outras características.

È mérito comentar que o papel de geneticistas tem sido reconhecido pela sociedade. Tanto é assim que geneticistas já foram laureados com o prêmio Nobel. Entre eles merece atenção o Dr. Norman Borlaug, que recebeu prêmio Nobel da Paz de 1970 pelas suas pesquisas, durante 25 anos com o melhoramento do trigo , e também a Dra. Bárbara McClintock, que trabalhando com citogenética do milho, recebeu prêmio Nobel de Fisiologia e Medicina em 1973, dado que seus trabalhos abriram perspectivas, entre outras coisas, para que muitas doenças hereditárias pudessem ser mais eficientemente controladas

1.2 NOÇÕES BÁSICAS PARA INTRODUÇÃO À GENÉTICA

Vamos agora enunciar separadamente os principais conceitos envolvidos no estudo da genética. As explanações abaixo serão breves, e aqueles interessados em se aprofundar no assunto podem encontrar em [7]. Existem também diversas fontes online, com diferentes níveis de profundidade. Algumas fontes interessantes são [2, 3, 11], para temas gerais de genética e [1, 6, 8–10, 12] para doenças de origem genética. Dentro destas fontes existem ainda links complementares.

O DNA

O **DNA**, ou Ácido Desoxirribonucleico, é um extenso composto orgânico, formado por moléculas que carregam informações responsáveis pelo funcionamento e desenvolvimento de organismos vivos.

Quimicamente, o DNA é um polímero formado de moléculas menores, chamadas **nucleotídeos**. Os nucleotídeos, por sua vez, são formados por

- Um fosfato (HPO_4);
- Uma desoxirribose (um tipo de açúcar com 5 carbonos)
- Uma base nitrogenada, que podem ser de 4 tipos: Adenina (A), Timina (T), Citosina (C) e Guanina (G).

Estruturalmente, o DNA é formado por dois longos filamentos, enrolados juntos em forma de espiral, também conhecida como dupla hélice.

A conexão entre os filamentos é feita por pontes de hidrogênio, formadas entre os pares de bases: A-T e C-G.

Esta longa sequência de bases nitrogenadas constitui a informação genética de um ser vivo. Tais sequências são traduzidas pelo organismo em proteínas, que regulam todo o funcionamento e desenvolvimento do corpo.

Em organismos como animais, plantas e fungos (os chamados procariontes) o DNA se encontra no núcleo das células organizado em estruturas chamados cromossomos, enquanto nos eucariontes (bactérias, por exemplo) ele se encontra disperso no citoplasma.

CROMOSSOMOS

Os **genes** cromossomos são estruturas moleculares formadas por cadeias de DNA e proteínas. Eles são a primeira grande componente de organização do material genético em um ser vivo. Os cromossomos são formados de partes menores, chamados de **genes**.

Em algumas células do organismo, conhecidas como **diplóides**, os cromossomos são organizados aos pares. Os dois cromossomos de um par são compostos pelos mesmos genes, e são conhecidos por **cromossomos homólogos**. Tais pares regulam as mesmas funções dentro do organismo.

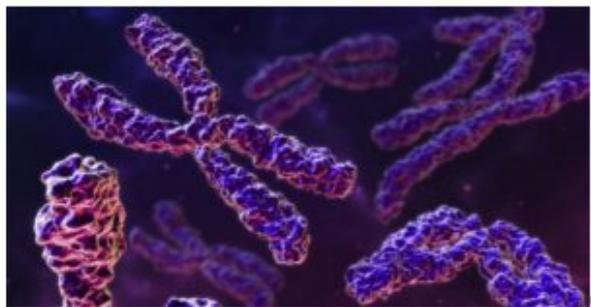


Figura 2: Imagem de cromossomos no núcleo de uma célula.

Em células **haplóides** (como espermatozóides e óvulos) os cromossomos não aparecem em pares, e não possuindo cromossomos homólogos.

A espécie humana possui 46 cromossomos, organizados em 23 pares de cromossomos homólogos. Deste total 23 cromossomos tiveram origem da mãe e os demais 23 do pai. Deste total, 44 cromossomos (22 pares) são denominados autossômicos, e outros 2 são cromossomos sexuais, relacionados à determinação do gênero.

Anomalias cromossômicas podem aparecer por erros durante a segregação celular, ocasionando o surgimento de síndromes genéticas, tais como alteração do número, tamanho ou do ordenamento de partes dos cromossomos.

CÉLULAS HAPLÓIDES E DIPLÓIDES

Em indivíduos eucariontes (animais, plantas, etc), as células podem ser classificadas de acordo com o material genético que carregam e sua função no organismo.

Células **diplóides**, como comentamos acima, são caracterizadas por apresentarem um número par de cromossomos em seu núcleo, organizados em pares de cromossomos homólogos. Estas são as células **somáticas**, responsáveis pela formação de órgãos e tecidos em seres multicelulares. São estas as células responsáveis pela síntese proteica que regula o organismo.

As células **haplóides**, por outro lado, são caracterizadas por possuírem metade dos cromossomos de uma célula diplóide, sem pares de cromossomos homólogos. Estas são as células com capacidades reprodutivas (gametas).

GENES

Presentes nos cromossomos, os genes são moléculas segmentares de DNA, que codificam a informação necessária para a síntese de uma proteína. São eles, portanto, os responsáveis pela expressão das características de um ser vivo. Cada característica, desde cor do olho, cor do cabelo, formato do nariz, até doenças de origem genética, são determinados por um ou mais genes em um cromossomo.

O **loco gênico** ou **locus** é a região onde este gene é encontrado dentro de um cromossomo. Nestas regiões situam-se os genes responsáveis por determinadas características. Situações em que há deslocamento dos genes do locus ao qual pertencem provocam alguns tipos de anomalias cromossômicas.

Um gene não aparece sempre da mesma forma, podendo apresentar formas diferentes, dadas por pequenas modificações na sequência de DNA, causadas por mutações.

Formas distintas de uma mesmo gene são conhecidas como **alelos**.

Cromossomos homólogos possuem a mesma sequência de genes, mas podem possuir diferentes alelos.

ALELOS DOMINANTES E RECESSIVOS

Alelos distintos de um mesmo gene atuam em conjunto no organismo, informando a formação de proteínas de mesma com mesma função. Por vezes, a ação de um dado alelo pode mascarar ou até mesmo sobrepor a expressão de outros alelos. Tais alelos são conhecidos como **dominantes**, enquanto os demais são conhecidos como **recessivos**.

Alguns pares de cromossomos, em um certo locus, apresentam pares de genes (alelos) similares (AA/aa). Tais locus são conhecidos como **homozigotos**. Se o locus é formado de aelos distintos (Aa), dizemos ser **heterozigoto**.

Os genes dominantes, como olhos castanhos, por exemplo, são geralmente representados sobre a letras maiúsculas (A). Enquanto isso, um gene recessivo (olhos claros, por exemplo) é expresso por letras minúsculas (a).

FENÓTIPO E GENÓTIPO

Fenótipo e Genótipo são conceitos bem abrangentes.

Para nós, e de modo simplificado, o **genótipo** é o conjunto de informações carregado pelos genes. Representam, portanto, o conjunto de informações carregadas pelos cromossomos para formação de um indivíduo.

Enquanto o genótipo especifica a constituição de genes de um determinado ser, o **fenótipo** o conjunto de características visuais observadas nos indivíduos. Em outras palavras, o modo como o genes se expressam.

Na determinação da cor de olho de um ser humano, por exemplo, dois indivíduos podem possuir a mesma cor de olho (fenótipo), mas informações genéticas distintas (genótipo). Um deles pode ser homocigoto dominante (AA) e o outro heterocigoto (Aa).

HERANÇA LIGADA AO SEXO

Como comentamos anteriormente, no ser humano, um par dos 46 cromossomos que carregamos são relacionados à características de gênero.

No ser humano e outros mamíferos, existem dois tipos de cromossomos sexuais, identificados por X e Y. Enquanto as mulheres apresentam um par de cromossomos X (XX), os homens possuem apenas um X, acompanhado de um Y (XY). Com isso, o gameta masculino será o determinante para o gênero biológico dos filhos

1.3 GENÉTICA E SEXO

Há dois tipos de animais e plantas com reprodução sexuada:

- (1) monóicos, em que cada indivíduo produz dois tipos de gametas, masculino e feminino;
- (2) dióicos, em que cada indivíduo produz apenas um tipo de gameta, feminino ou masculino.

Em organismos dióicos, a diferença sexual primária diz respeito ao tipo de gametas e aos órgãos sexuais primários pelos quais são produzidos. Cada sexo também exibe muitos caracteres sexuais secundários. Em seres humanos estes incluem voz, distribuição de gordura e cabelos no corpo e detalhes de estrutura de musculatura e esqueleto; em *Drosophila*, mosca das frutas, incluem o número de segmentos abdominais e assim por diante.

Em espécies monóicas, a reprodução é feita através do encontro do material genético de dois indivíduos distintos, ou até mesmo por troca de material em um mesmo indivíduo.

No maracujá, por exemplo, cada flor possui órgãos reprodutores femininos e masculinos, mas são incapazes de se auto-fecundar. Para acontecer a reprodução algum agente externo, como abelhas, carregam o material genético masculino (na forma de pólen) involuntariamente, e depositam no estigma (orgão sexual feminino), quando visitam outra flor.

1.3.1 Cromossomos sexuais e determinação genética do sexo

Um par de cromossomos citologicamente distintos proporciona a base para a determinação do sexo na maioria dos animais superiores e um número reduzido de plantas. Tais cromossomos são conhecidos como cromossomos sexuais, enquanto os demais recebem a denominação de autossomos.

Existem basicamente três sistemas de determinação do sexo. XY, XO e ZW.

XX-XO

Em 1891, um biólogo alemão chamado H. Henking, notou que parte dos espermatozoides gerados em certas espécies de insetos continha uma estrutura nuclear extra, que deu o nome de o “corpo X” (por ser um elemento estranho presente na célula).

O significado dessa estrutura não foi imediatamente entendido, mas em 1902 Clarence McClung, um biólogo norte-americano, relatou que as células somáticas do gafanhoto fêmea continham 24 cromossomos, enquanto as do macho tinham apenas 23.

Três anos depois, Edward B. Wilson e N. M. Stevens conseguiram seguir a ovulogênese e a espermatogênese em vários insetos, e perceberam que o corpo X era de fato um cromossomo, e assim ele passou a ser conhecido como o cromossomo X.

Assim, em muitos insetos há uma diferença cromossômica entre os sexos, sendo as fêmeas referidas como XX (com dois cromossomos X) e os machos como XO (com apenas um cromossomo X, e nenhum outro cromossomo sexual).

Em consequência da meiose, todos os óvulos de tais espécies carregam um cromossomo X, mas apenas metade dos espermatozoides tem um, e a outra metade nenhum.

Sistema XX-XY

O chamado tipo XY de diferenciação sexual ocorre em uma grande variedade de animais, incluindo a *Drosophila*, os mamíferos (como o homem) e em alguns vegetais (por exemplo, os do gênero *Lychnis* de angiospermas). Em tais casos as fêmeas são também XX, mas os machos têm, além do X único, um cromossomo de tamanho diferente; que foi chamado cromossomo Y. Assim, nestas espécies, os machos são XY.

Uma característica importante dos cromossomos X, é que estes apresentam mais genes que Y, e portanto alguns genes presentes em X, não possuem alelos correspondentes em Y.

Da mesma forma alguns locus de Y, não possuem similares em X, formando características exclusivas do gênero masculino.

Dessa maneira, podemos determinar heranças genéticas devidas ao gênero. Um gene existente em X, mas sem correspondente em Y, determina uma característica que o indivíduo do sexo masculino, por exemplo, só pode herdar da mãe.

Se, por outro lado, a característica for exclusiva de Y, apenas os descendentes do gênero masculino à herdarão, e sempre do pai. Tais heranças são menos interessantes em um contexto de modelagem, uma vez que o cromossomo Y de um indivíduo deve ser sempre igual ao do pai.

Na formação dos gametas, aproximadamente metade dos espermatozoides carrega o cromossomo X e metade o Y. Já os óvulos sempre carregam o X.

ZZ-ZW

O último dos tipos principais de diferença cromossômica entre os sexos é aquela em que a fêmea é heterogamética (ZW) e o macho homogamético (ZZ). Frequentemente designam-se os cromossomos sexuais nesse caso como Z e W para evitar confusão com os casos em que a fêmea seja homogamética. Aves (incluindo as domésticas), borboletas, mariposas e alguns peixes pertencem a este grupo.

1.3.2 Herança ligada aos cromossomos sexuais

Conhecem-se mais de 150 caracteres confirmados ou altamente prováveis de serem ligados ao cromossomo X; a maioria deles recessiva. O primeiro a ser descrito na literatura, um tipo de daltonismo verde-vermelho em que faltam cones sensíveis ao verde (daltonismo deutan), é devido a um alelo recessivo ligado ao X.

Neste tipo de daltonismo, as mulheres, é claro, podem ser homozigotas normais, heterozigotas ou (raramente) homozigotas com alelo defeituoso, pelo simples fato de possuir dois cromossomos X. Com isso, possuem uma probabilidade maior de receber um alelo para visão normal de pelo menos um dos pais. As mulheres heterozigotas variam no grau em que é afetada sua visão de cores, dependendo da proporção de suas células retiniais que expressam o alelo normal. Os homens, por outro lado, recebem ou um alelo dominante para normal ou um alelo recessivo para visão defeituosa para cores vermelho-verde (da mãe, que contribuiu com seu cromossomo X). O daltonismo deutan parece ser o caráter ligado ao sexo mais comumente encontrado em seres humanos.

A hemofilia, um distúrbio bem conhecido em que ocorre uma deficiência na coagulação do sangue, devido à falta de substrato necessária para atividade tromboplastínica, é também uma condição recessiva ligada ao sexo. São reconhecidos dois tipos de hemofilia ligada ao sexo:

1. Hemofilia A, caracterizada pela falta de globulina anti-hemofílica (Fator VIII).
2. Hemofilia B, ou “doença de Christmas” (nome da família em que primeiro foi descrita com detalhes), que resulta de um defeito no componente tromboplastínica do plasma (PTC ou Fator IX).

Esta última sendo uma forma mais suave da condição. Conhecem-se muitas variantes desse alelo recessivo, cada um afetando uma porção diferente da cadeia polipeptídica da enzima. Essas variantes têm efeitos que vão de grave a suave a nenhuma deficiência da enzima; algumas até produzem atividade enzimática aumentada.

Outra classe de caracteres ligados ao sexo refere-se aos genes situados no cromossomo Y, relacionados à expressão sexual masculina e passadas somente do pai para o filho.

Foi sugerido provisoriamente (ainda não confirmado) que o(s) gene(s) que controla(m) a espermatogênese estejam no braço longo do Y. Pensa-se também que alguns dos genes que contribuem para a altura também estejam nesse cromossomo, assim como os genes que determinam maturação mais lenta do indivíduo.

HERANÇA DE GENES INFLUENCIADOS PELO SEXO

Em contraste com os genes limitados ao sexo, em que a expressão de um caráter é limitada a um sexo, os genes influenciados pelo sexo são aqueles cuja dominância é influenciada pelo sexo do portador. É no comportamento dos heterozigotos que detectamos essa influência, pois nos homozigotos, o alelo se manifesta como deveria fazê-lo em qualquer sexo. Mas, no heterozigoto, ele age como dominante num sexo e recessivo no outro. Um exemplo interessante é a calvície .

Apesar da calvície poder aparecer por qualquer de várias causas (por exemplo, doenças, radiação), a calvície “padronizada” exibe um definido padrão genético. Nesta condição, o cabelo rarefaz-se gradualmente no topo, levando finalmente a uma faixa de cabelo na parte baixa da cabeça. A calvície padrão é mais prevalente em homens do que em mulheres, em que é rara e geralmente envolve pronunciado afinamento do cabelo em vez de perda total no topo da cabeça. A calvície padrão é devida a um par de alelos autossômicos .

Exemplo 1.1.

Calvície

A calvície pode ter origem em vários fatores: seborréia, sífilis, distúrbio da tiróide etc. Entretanto a intensidade e alguns tipos de calvície pode ser hereditária. A calvície hereditária tem o seguinte mecanismo genético:

Genótipo	Homem	Mulher
BB	Calvo	Calva
Bb	Calvo	Não calva
Bb	Não calvo	Não calva

Neste caso, o alelo B determina a calvície e o alelo b determina a persistência de cabelos. O alelo B domina b, quando em um genótipo do indivíduo do sexo masculino. Em indivíduos do sexo feminino o alelo b domina o alelo B.

Pelagem de gado ayrshire

Nesta raça de gado leiteiro o animal branco pode apresentar áreas (manchas) no pescoço que podem ter coloração vermelha ou castanha. Os seguintes alelos determinam a herança deste caráter:

- M1 = castanho
- M2 = vermelho

Sendo que M1 domina M2 em organismos do sexo masculino, mas é dominado por M2 nos indivíduos do sexo feminino. Assim, são verificados os seguinte fenótipos associados aos genótipos atribuídos ao gene M1/M2:

Genótipo	Macho	Fêmea
M1 M1	Castanho	Castanho
M1 M2	Castanho	Vermelho
M2 M2	Vermelho	Vermelho

Presença de chifres em carneiros

Com relação a este caráter temos o seguinte mecanismo genético:

- H1 = com chifre
- H2 = sem chifre

Neste caso, H1 domina H2 nos machos e H2 domina H1 nas fêmeas, conforme ilustrado no quadro a seguir:

Genótipo	Macho	Fêmea
H1 H1	Com chifre	Com chifre
H1 H2	Com chifre	Sem chifre
H2 H2	Sem chifre	Sem chifre

Algumas doenças ligadas ao cromossomo

As doenças abaixo podem ser ligadas a quantidade de cromossomos:

- A síndrome de Klinefelter.

Um em cerca de 500 nascimentos “masculinos” produz um indivíduo com um conjunto particular de anormalidades conhecido coletivamente como a síndrome de Klinefelter. Essas pessoas têm um fenótipo geral masculino; a genitália externa é essencialmente de morfologia normal. Embora haja alguma variabilidade em outras características, os testículos são tipicamente pequenos, geralmente não produzem espermatozoides e a inteligência pode ser normal ou levemente retardada. Os braços são mais longos do que a média, é comum algum grau de desenvolvimento de seios e a voz tende a ser de timbre mais alto do que em homens normais. Os indivíduos Klinefelter são cromatina sexual-positivos; o cariótipo o mostra 47 cromossomos, isto é, 47, XXY.

Os indivíduos XXY podem resultar de fertilização de um óvulo XX por um espermatozóide Y ou de um óvulo X por um espermatozóide XY.

- A síndrome de Turner.

Uma segunda anomalia importante dos cromossomos sexuais de interesse aqui é a síndrome de Turner, na qual o indivíduo apresenta um fenótipo geral feminino, mas com certos desvios típicos. Estes incluem baixa estatura, pescoço “alado”, inserção baixa dos cabelos na nuca e tórax largo em forma de escudo. Frequentemente nota-se leve retardamento mental, mas algumas Turner atingem resultados elevados em testes padronizados de QI.

Os caracteres sexuais secundários não se desenvolvem; o desenvolvimento das mamas é de ausente a muito pequeno, os pêlos pubianos são reduzidos ou ausentes e os pêlos axilares não se desenvolvem. As genitálias permanecem essencialmente infantis.

O problema é geralmente causado pela ausência de um cromossomo X, ou por defeitos estruturais neste cromossomo. Geralmente as Turner são estéreis, embora pelo menos um nascimento normal e várias gestações tenham sido relatados em prováveis Turner

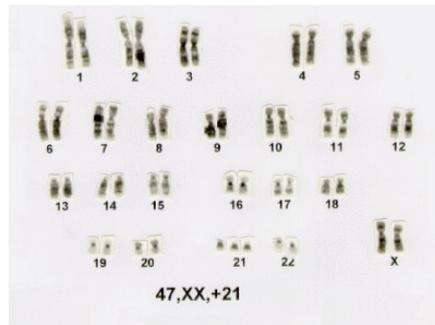


Figura 3: Cromossomos de um indivíduo com Síndrome de Down

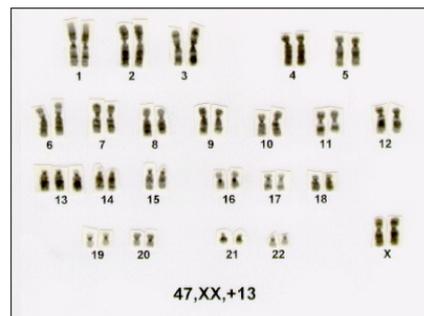


Figura 4: Cromossomos de um indivíduo com Síndrome de Down

- Síndrome de Down

Há um erro na contagem de cromossomos, ao invés do indivíduo ter 46 cromossomos, ele apresenta 47. O par do cromossomo 21 recebeu mais um cromossomo, ficando com um trio de cromossomos.

- Síndrome de Patau

Igualmente a Síndrome de Down, há um erro na contagem de cromossomos, ao invés do indivíduo ter 46 cromossomos, ele apresenta 47. O par do cromossomo 13 recebeu mais um cromossomo, ficando com um trio de cromossomos, sendo trissômico.

- Síndrome de Edward

Igualmente a Síndrome de Down, há um erro na contagem de cromossomos, ao invés do indivíduo ter 46 cromossomos, ele apresenta 47. O par do cromossomo 18 recebeu mais um cromossomo, ficando com um trio de cromossomos, sendo trissômico.

Outras doenças podem ser ligadas apenas aos cromossomos sexuais X ou Y. Quando o gene alterado está no cromossomo X, este tipo de herança tem o padrão recessivo; é a herança materna. Neste caso, os filhos homem herdam genes do cromossomo X apenas da mãe. Enquanto as filhas mulheres herdam um do pai e outro da mãe. As manifestações vão estar presentes nos machos pois apresentam apenas um cromossomo X, ou seja, não apresentam nenhum gene normal para aquela característica.

- Algumas doenças relacionadas ao cromossomo X: daltonismo, hemofilia, calvície (homozigoto recessivo).
- Algumas doenças relacionadas ao cromossomo Y: hipertricos, calvície (apenas um alelo dominante)

FERRAMENTAS MATEMÁTICAS

Acredita-se que um dos motivos das ideias de Mendel ficarem incompreendidas por quase 3 décadas foi a matemática envolvida. Ele partiu do princípio que a formação de gametas seguia princípios aleatórios, especialmente na distribuição de fatores, e modelos aleatórios ainda não era exatamente comuns, especialmente nas ciências biológicas.

A ideia era simples: a formação de um determinado tipo de gameta, com um alelo oriundo de um par de genes, é um evento aleatório assim como tirar uma carta de um baralho. Assim, um indivíduo heterozigoto (Aa) teria a mesma probabilidade de formar gametas A ou gametas a . Seguindo estas ideias, neste modelo um casal teria sempre 50% de probabilidade ter um menino e 50% de ter uma menina, por exemplo.

Tais ideias ainda podem ser estendidas, inserindo aí fatores externos ou até mesmo temporais. Neste trabalho exploraremos modelos de propagação de genes, que adiciona às ideias de Mendel um componente temporal.

Nas próximas seções exploraremos rapidamente os conceitos matemáticos necessários para entendermos os modelos.

Começaremos falando das equações de diferença, que cuida dos elementos temporais do modelo, e nos será muito útil na sua análise. Aqui faremos uma breve apresentação, tratando principalmente de casos que serão de interesse para este trabalho. Para maiores detalhes sobre o assunto sugerimos [4].

Em seguida falaremos brevemente da teoria das probabilidades, que serve de base para o modelo. Aqui os comentários serão ainda menos profundos, e novamente apresentaremos

apenas conceitos importantes para nós. A teoria completa é vasta, e pode ser vista em maiores detalhes em [13, 14].

2.1 EQUAÇÕES DE DIFERENÇA

Suponha que queremos estudar o evolução do tamanho $p(n)$ de uma população ao longo de um período de tempo. Fazer hipóteses sobre o comportamento de longo prazo da função $p(n)$ parece complicado e arriscado. Além disso, é exatamente o que esperamos obter com o modelo. De modo que estaríamos “começando pelo fim”.

No entanto, podemos tentar fazer hipóteses muito mais confiáveis, se nos concentrarmos na relação entre as populações em instantes próximos de tempo.

Podemos supor, por exemplo, que os membros de uma dada população se reproduzem, gerando descendentes à uma taxa de λ indivíduos por unidade de tempo, onde $\lambda > 0$. Isso seria representado pela equação $p(n + 1) = \lambda \cdot p(n)$.

Podemos incluir fatores ambientais, e dizer que quando a população é muito grande, a taxa de reprodução reduz. Com isso a taxa λ dependeria da população p , e podemos fazer que $\lambda(p) := r(K - p)$, onde K é uma espécie de tamanho máximo de população suportada pelo ambiente, e $r > 0$. Com isso, teríamos um modelo do tipo

$$p(n + 1) = rp(n)(K - p(n)).$$

Este tipo de equações, bastante comum em modelos de evolução temporal, são conhecidos como equações de diferença. Neles tentamos descrever uma sequência $p(n)$ pelas relações entre valores próximos de n . O nome vem da versão contínua, onde este tipo de equação recebe o nome de equação diferencial.

A seguir vamos estudar algumas das equações mais comuns, mostrando como encontrar soluções gerais para elas.

2.1.1 Sequências definidas por recorrência

Neste trabalho nos concentraremos em um tipo específico de equações de diferença: as sequências definidas por relações de recorrência temporal.

Informalmente, são sequências definidas por regras que permitem calcular qualquer termo em função dos anteriores (usualmente do antecessor imediato ou de uma quantidade pequena de antecessores imediatos).

Mais formalmente, tratam-se de sequências $(x_n)_{n \geq n_0} \subset \mathbb{R}$ onde

$$x_n = f(x_{n-1}, x_{n-2}, \dots, x_{n-k}),$$

para alguma função $f : \mathbb{R}^k \rightarrow \mathbb{R}$.

O valor k acima, que representa até onde no passado devemos voltar para encontrar o valor de um dado termo da sequência é conhecido como ordem da relação.

Assim, uma relação de primeira ordem tem a forma $x_n = f(x_{n-1})$, e o valor de um termo depende apenas do valor termo anterior. Já uma relação de segunda ordem tem a forma $x_n = f(x_{n-1}, x_{n-2})$.

Exemplo 2.1 (PA's e PG's). Uma **progressão aritmética** (PA) de primeiro termo $a \in \mathbb{R}$ e a razão $r \in \mathbb{R}$ e uma sequência definida por $x_1 = a$ e

$$x_{n+1} = x_n + r,$$

para $n \geq 1$.

Uma **progressão geométrica** (PG) de primeiro termo $a \in \mathbb{R}$ e a razão $r \in \mathbb{R}$ e uma sequência definida por $x_1 = a$ e

$$x_{n+1} = r \cdot x_n,$$

para $n \geq 1$.

Ambas são equações de primeira ordem.

Exemplo 2.2 (Sequência de Fibonacci). A chamada sequência de *Fibonacci* é a sequência definida dada por $x_1 = 1$, $x_2 = 1$, e

$$x_{n+2} = x_n + x_{n+1},$$

para $n \geq 1$.

Ou, equivalentemente,

$$x_n = x_{n-2} + x_{n-1},$$

para $n \geq 3$.

A equação que gera a sequência de Fibonacci é uma equação de segunda ordem.

Este exemplos nos mostram algo simples, mas importante: se quisermos encontrar uma solução para uma equação deste tipo não basta a lei de recorrência; é preciso também fornecer o(s) primeiro(s) termo(s) da sequência.

No caso de uma PA dada por $x_{n+1} = x_n + 2$, se definirmos $x_1 = 1$ encontramos a sequência $\{1, 3, 5, 7, \dots\}$ dos números ímpares. Se, por outro lado, escolhermos $x_1 = 2$ encontramos a sequência $\{2, 4, 6, 8, \dots\}$ de números pares.

A mesma brincadeira com a regra de formação da sequência de Fibonacci, mostra que são necessários os dois primeiros termos da sequência, sem os quais não conseguimos calcular o terceiro, e assim o resto da sequência.

Exemplo 2.3. Quantas são as sequências de 10 termos, todos pertencentes a $0, 1, 2$, que não têm dois termos consecutivos iguais a 0?

Seja x_n o número de sequências com n termos respeitando as condições do enunciado. Vamos contar separadamente as sequências, de acordo com o seu termo inicial.

- Se fixarmos o primeiro termo da sequência como sendo 1, estamos livres para escolher os próximos $n - 1$ termos, e assim o total de sequências começando por 1 é x_{n-1} ;
- Pela mesma razão, o total de sequências começando por 2 também é x_{n-1} ;
- Se fixarmos o primeiro termo como 0, limitamos escolha do segundo termo, que só pode assumir os valores 1 ou 2. Assim, como para cada escolha de segundo termo temos x_{n-2} formas de escolher os termos restantes, o total de sequências começando por 0 é $2x_{n-2}$.

Logo $x_n = 2x_{n-1} + 2x_{n-2}$, para $n \geq 3$.

Vale também que $x_1 = 3$. Para encontrar x_2 , basta perceber que a única sequência de 2 termos que quebra a regra é a sequência $(0,0)$. Assim, como temos 9 sequências disponíveis, encontramos $x_2 = 8$.

Daí, obtemos

$$x_1 = 3$$

$$x_2 = 8$$

$$x_3 = 2 \cdot 3 + 2 \cdot 8 = 22$$

$$x_4 = 2 \cdot 8 + 2 \cdot 22 = 60$$

$$x_5 = 2 \cdot 22 + 2 \cdot 60 = 164$$

$$x_6 = 2 \cdot 60 + 2 \cdot 164 = 448$$

$$x_7 = 2 \cdot 164 + 2 \cdot 448 = 1224$$

$$x_8 = 2 \cdot 448 + 2 \cdot 1224 = 3344$$

$$x_9 = 2 \cdot 1224 + 2 \cdot 3344 = 9136$$

$$x_{10} = 2 \cdot 3344 + 2 \cdot 9136 = 24960.$$

E assim, $x_{10} = 24960$.

2.1.2 *Recorrências Lineares de Primeira Ordem*

Começaremos agora a tratar um caso especial de equações de recorrência: as chamadas recorrências lineares. Esta é uma classe grande de equações, importante no estudo de diversos modelos, e para as quais as soluções explícitas são mais fáceis de se calcular.

Neste texto trataremos apenas nas equações lineares de primeira e segunda ordem.

Uma equação de recorrência de primeira ordem é dita linear se possui a forma

$$x_n = g(n)x_{n-1} + h(n),$$

onde g e h são funções reais, dependendo apenas de n .

Se $h(n) \equiv 0$ diremos que a equação é homogênea.

As equações $x_{n+1} = 2x_n - n^2$ e $x_{n+1} = nx_n$ são lineares, sendo que a última destas é homogênea. As recorrências $x_{n+1} = x_n^2$ e $x_n = e^{x_{n-1}}$ são não-lineares.

Equações Homogêneas

Começemos com um exemplo.

Exemplo 2.4. Vamos encontrar a solução da equação $x_{n+1} = n \cdot x_n$

$$\begin{aligned}x_2 &= 1 \cdot x_1 \\x_3 &= 2 \cdot x_2 \\x_4 &= 3 \cdot x_3 \\&\vdots \\x_n &= (n-1) \cdot x_{n-1}\end{aligned}$$

Multiplicando os termos de todas as igualdades obtemos

$$x_2 \cdot x_3 \cdots x_n = 1 \cdot 2 \cdots (n-1) \cdot x_1 \cdot x_2 \cdots x_{n-1},$$

e portanto temos

$$x_n = (n-1)! x_1$$

Com isso, obtemos a solução geral: $x_n = C \cdot (n-1)!$, onde a constante C é exatamente o primeiro termo da sequência.

A solução do exemplo acima pode claramente ser generalizada. Assim, se queremos resolver a equação

$$x_{n+1} = g(n)x_n,$$

basta fazer

$$\begin{aligned}x_2 &= g(1) \cdot x_1 \\x_3 &= g(2) \cdot x_2 \\x_4 &= g(3) \cdot x_3 \\&\vdots \\x_n &= g(n-1) \cdot x_{n-1}\end{aligned}$$

Procedendo como no exemplo, encontramos

$$x_n = C \cdot g(1) \cdot g(2) \cdots g(n-1), \quad (2.1)$$

onde $C = x_1$, é o primeiro termo da sequência.

Equações da Forma $x_{n+1} = x_n + f(n)$

Para equações deste tipo podemos proceder de maneira similar às equações homogêneas, nos aproveitando do fato do termo que acompanha x_{n-1} ser igual a 1.

Para isso fazemos

$$\begin{aligned} x_2 &= x_1 + f(1) \\ x_3 &= x_2 + f(2) \\ x_4 &= x_3 + f(3) \\ &\vdots \\ x_n &= x_{n-1} + f(n-1) \end{aligned}$$

E somando todas as equações obtemos

$$x_2 + \cdots + x_n = x_1 + x_2 + \cdots + x_{n-1} + f(1) + \cdots + f(n-1),$$

de onde obtemos

$$x_n = C + \sum_{k=1}^{n-1} f(k), \quad (2.2)$$

onde, mais uma vez, temos $C = x_1$.

Exemplo 2.5. Para resolver a recorrência $x_{n+1} = x_n + 2^n$, com $x_1 = 1$, podemos seguir os passos acima e fazer

$$\begin{aligned} x_2 &= x_1 + 2 \\ x_3 &= x_2 + 2^2 \\ x_4 &= x_3 + 2^3 \\ &\vdots \\ x_n &= x_{n-1} + 2^{n-1} \end{aligned}$$

E somando, obter

$$\begin{aligned} x_n &= x_1 + (2 + 2^2 + 2^3 + \dots + 2^{n-1}) \\ &= 1 + 2 + 2^2 + 2^3 + \dots + 2^{n-1} \\ &= \frac{2^n - 1}{2 - 1} \\ &= 2^n - 1. \end{aligned}$$

Equações não homogêneas em geral

Vamos agora mostrar como encontrar a solução geral de uma equação linear de primeira ordem qualquer. Ou seja, uma equação do tipo

$$x_{n+1} = g(n)x_n + f(n). \quad (2.3)$$

Até este momento tratamos apenas dos casos particulares $g(n) \equiv 1$ e $f(n) \equiv 0$, e seria interessante encontrar um modo de usar tais soluções para resolver nosso problema.

Usar a soma ou produto das equações para diferentes valores de n , como fizemos nos casos anteriores, não funciona, uma vez que não seria possível cancelar os termos x_k com $k = 2, \dots, n - 1$, como gostaríamos. Mas isso ainda não significa que a solução das equações anteriores não nos seriam úteis.

Um modo bastante comum de começar este tipo de problema é tomar a solução do problema simplificado e tentar entender o que falta nesta solução para encontrarmos uma solução para o problema que temos.

Talvez seja interessante ilustrarmos essa ideia com um exemplo.

Exemplo 2.6. Considere a equação $x_{n+1} = \frac{1}{2}x_n + 1$.

Ainda não sabemos como resolver esta equação, mas estudamos como resolver a equação homogênea $a_{n+1} = \frac{1}{2}a_n$, que tem como solução

$$a_n = C \cdot \frac{1}{2^{n-1}}.$$

Para simplificar, escolha $C = 1$, fazendo

$$a_n = \frac{1}{2^{n-1}}.$$

A ideia é estudar o que falta nesta sequência para transformá-la em uma solução da equação original, e temos ao menos dois modos de fazer isso. Podemos somar algo à solução encontrada, ou multiplicar a sequência a_n por alguma coisa. Ou seja, podemos buscar soluções do tipo

$$x_n = a_n + z_n \quad \text{OU} \quad x_n = a_n \cdot y_n.$$

Vamos tentar a primeira opção. Substituindo na equação original temos

$$a_{n+1} + z_{n+1} = \frac{1}{2}(a_n + z_n) + 1 = \frac{1}{2}a_n + \frac{1}{2}z_n + 1,$$

e como $a_{n+1} = \frac{1}{2}a_n$, encontramos

$$z_{n+1} = \frac{1}{2}z_n + 1.$$

Assim z_n deve ser solução da mesma equação que queremos resolver! Apesar de interessante, esta informação nos ajuda muito neste momento.

Tentemos então a segunda opção.

$$a_{n+1} \cdot y_{n+1} = \frac{1}{2}a_n \cdot y_n + 1,$$

e como $a_{n+1} = \frac{1}{2}a_n$, encontramos

$$y_{n+1} = y_n + \frac{1}{\frac{1}{2}a_n},$$

e portanto

$$y_{n+1} = y_n + 2^n,$$

que sabemos exatamente como resolver.

Do exemplo 2.5 sabemos que

$$y_n = y_1 + 2 + \cdots + 2^{n-1} = y_1 + 2^n - 2,$$

e portanto

$$x_n = a_n \cdot y_n = \frac{1}{2^{n-1}} \cdot y_1 + 2 - \frac{1}{2^{n-2}} = \frac{1}{2^n}(2y_1 - 4) + 2.$$

Como

$$y_1 = \frac{x_1}{a_1} = x_1,$$

concluimos que

$$x_n = 2^{-n}(2x_1 - 4) + 2.$$

Entendemos assim uma possível forma de encontrar a solução de uma equação do tipo (2.3).

Se a_n é uma solução não-nula da equação homogênea $x_{n+1} = g(n)x_n$, então a substituição

$$x_n = a_n y_n$$

deve transformar a equação $x_{n+1} = g(n)x_n + f(n)$ algo da forma $y_{n+1} = y_n + h(n)$.

Vejam os. A substituição de $x_n = a_n y_n$ em (2.3) transforma

$$x_{n+1} = g(n)x_n + f(n) \quad \text{em} \quad a_{n+1}y_{n+1} = g(n)a_n y_n + f(n).$$

Mas, $a_{n+1} = g(n)a_n$, pois a_n é solução de $x_{n+1} = g(n)x_n$, e portanto a equação pode ser escrita como

$$g(n)a_n y_{n+1} = g(n)a_n y_n + f(n). \quad (2.4)$$

Ou seja,

$$y_{n+1} = y_n + \frac{f(n)}{g(n)a_n}.$$

Exemplo 2.7. Pode-se por exemplo resolver a equações $x_{n+1} = \frac{n+1}{n}x_n + (n+1)2^n$, $x_1 = 2$.

Começamos por resolver a equação homogênea $a_{n+1} = \frac{n}{n+1}a_n$. Sabemos que, neste caso

$$a_n = a_1 \cdot \frac{2}{1} \cdot \frac{3}{2} \cdots \frac{n-1}{n-2} \cdot \frac{n}{n-1} = a_1 n.$$

Podemos escolher então $a_1 = 1$ e tomar $a_n = n$.

Fazendo a substituição $x_n = n \cdot y_n$, obtemos

$$(n+1) \cdot y_{n+1} = \frac{n+1}{n}(n \cdot y_n) + (n+1)2^n.$$

Ou seja,

$$y_{n+1} = y_n + 2^n,$$

com $y_1 = x_1 = 2$.

Daí

$$\begin{aligned} y_n &= y_1 + 2^1 + 2^2 + \dots + 2^{n-1} \\ &= 2 + 2^1 + 2^2 + \dots + 2^{n-1} \quad . \\ &= 1 + 2^n - 1 \end{aligned}$$

E portanto $y_n = 2^n$.

Fizemos a substituição $x_n = n \cdot y_n$ e obtivemos a solução $y_n = 2^n$ para a nova equação.

Segue daí que

$$x_n = n \cdot 2^n.$$

2.1.3 Equações Lineares de Segunda Ordem

Trataremos agora de equações de segunda ordem que podem ser escritas como

$$x_{n+2} = p(n) \cdot x_{n+1} + q(n) \cdot x_n + f(n).$$

O aparecimento do termo de segunda ordem $q(n) \cdot x_n$ impõe uma dificuldade maior ao problema, e soluções gerais, como encontramos no caso de primeira ordem, ficam bem mais complicadas de serem geradas.

Para compreender melhor a fonte de tais dificuldades, vamos considerar o caso homogêneo ($f(n) \equiv 0$), e observar que

$$v_{n+1} = \begin{bmatrix} x_{n+2} \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} p(n) \cdot x_{n+1} + q(n) \cdot x_n \\ x_{n+1} \end{bmatrix} = \begin{bmatrix} p(n) & q(n) \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = M(n) \cdot v_n, \quad (2.5)$$

com

$$M(n) = \begin{bmatrix} p(n) & q(n) \\ 0 & 1 \end{bmatrix}.$$

Aplicando a mesma técnica da primeira ordem, encontramos

$$v_n = M(n-1) \cdot M(n-2) \cdots M(1) \cdot v_1.$$

A solução é similar, mas o produto matricial $M(n-2) \cdots M(1)$ é sensivelmente mais complicado para resolver. Mesmo no caso de matriz constante (não dependente de n), o cálculo das potências M^n usam técnicas que fogem aos objetivos deste trabalho.

Vamos nos concentrar então na solução de equações do tipo

$$x_{n+2} + px_{n+1} + qx_n = 0, \quad (2.6)$$

com $q \neq 0$ (se $q=0$, a recorrência é, na verdade, de primeira ordem), que são no fundo o único tipo de equação que precisaremos neste trabalho.

Começaremos, como já fizemos anteriormente, testando soluções de equações que conhecemos e, com sorte, chegaremos a algo que nos dê alguma direção.

Como vimos, equações de primeira ordem do tipo $x_{n+1} = r \cdot x_n$ tem soluções do tipo $x_n = C \cdot r^n$. Este parece um bom ponto para começarmos.

Vamos tentar soluções do tipo $x_n = C \cdot r^n$, e ver o que conseguimos.

Substituindo em (2.6) encontramos

$$C \cdot r^{n+2} + p \cdot C \cdot r^{n+1} + q \cdot C \cdot r^n = 0,$$

e dividindo por $C \cdot r^n$ encontramos

$$r^2 + pr + q = 0.$$

Concluimos assim que se r é raiz do polinômio acima, então $x_n = C \cdot r^n$ é uma solução da equação estudada. Temos assim um ponto de partida.

Definição 2.1. Dada uma equação linear homogênea do tipo

$$x_{n+2} + px_{n+1} + qx_n = 0,$$

chamaremos de **polinômio característico** da equação linear o polinômio

$$p(r) = r^2 + pr + q.$$

A equação

$$r^2 + pr + q = 0$$

é conhecida como **equação característica**.

Veremos a seguir que as raízes do polinômio característico desempenham um papel fundamental na expressão da solução geral para a recorrência.

Raízes do polinômio característico e soluções da equação linear

Antes de seguir para a solução das equações, vamos primeiro explorar duas propriedades importantes de suas soluções.

Tome então uma equação homogênea

$$x_{n+2} + p(n) \cdot x_{n+1} + q(n) \cdot x_n = 0,$$

e sejam y_n e z_n duas soluções distintas de tal equação.

Observe que se $x_n = c_1 y_n + c_2 z_n$ com $c_1, c_2 \in \mathbb{R}$, então

$$\begin{aligned} x_{n+2} + p(n)x_{n+1} + q(n)x_n &= (c_1 y_{n+2} + c_2 z_{n+2}) + p(c_1 y_{n+1} + c_2 z_{n+1}) + q(c_1 y_n + c_2 z_n) \\ &= c_1 (y_{n+2} + p y_{n+1} + q y_n) + c_2 (z_{n+2} + p z_{n+1} + q z_n) \\ &= c_1 \cdot 0 + c_2 \cdot 0 \\ &= 0, \end{aligned}$$

de onde segue que $x_n = c_1 y_n + c_2 z_n$ é uma nova solução da equação homogênea.

Esta propriedade é o que dá o nome de *linear* para a equação, e será essencial em toda análise que faremos a seguir.

A segunda propriedade vem de observar que se sabemos os dois termos iniciais x_1 e x_2 de uma equação de segunda ordem, conseguimos calcular de forma única o termo x_3 . Agora, usando x_2 e x_3 , definimos unicamente x_4 , e assim por diante.

Deste modo, vemos que dados $a, b \in \mathbb{R}$ existe no máximo uma solução x_n da equação linear tal que $x_1 = a$ e $x_2 = b$.

Tendo estas duas propriedades em mente, sigamos nossa análise.

Como vimos, ao buscar uma solução do tipo $C \cdot r^n$ para a equação $x_{n+2} + p \cdot x_{n+1} + q \cdot x_n = 0$ encontramos o seu polinômio característico, cujas raízes nos indicam um caminho para encontrar soluções da equação.

Como o polinômio característico tem grau 2, temos 3 casos para analisar: duas raízes reais, uma raíz real e nenhuma raíz.

O caso em que o polinômio não possui raízes parece não esclarecer nada sobre o problema, nos fazendo voltar à estaca zero. A verdade é que esta ausência de raízes traz informações importante sobre o problema, mas tais soluções não serão necessárias neste trabalho, além de envolver uma matemática que está além do que buscamos apresentar, e portanto deixaremos de lado.

A seguir analisaremos os dois outros casos separadamente.

RESOLVENDO A EQUAÇÃO: DUAS RAÍZES DISTINTAS

Se as raízes de $r^2 + pr + q = 0$ são r_1 e r_2 , com $r_1 \neq r_2$, então

$$y_n = r_1^n$$

e

$$z_n = r_2^n$$

são soluções da equação de segunda ordem.

É interessante notar que, como $q \neq 0$, as raízes são não nulas, e portanto as soluções são também não nulas.

Segue que

$$x_n = c_1 r_1^n + c_2 r_2^n,$$

c_1 e c_2 constantes formam uma família de soluções para a equação.

Tome então $a_1, a_2 \in \mathbb{R}$ e observe que

$$\begin{aligned} c_1 r_1 + c_2 r_2 &= a_1, \\ c_1 r_1^2 + c_2 r_2^2 &= a_2, \end{aligned} \tag{2.7}$$

tem solução única. Isso por que, como $r_1 \neq r_2$, então o determinante

$$\begin{vmatrix} r_1 & r_2 \\ r_1^2 & r_2^2 \end{vmatrix} = r_1 r_2^2 - r_1^2 r_2 = r_1 r_2 (r_2 - r_1) \neq 0.$$

Isso nos mostra que temos uma solução da forma $x_n = c_1 r_1^n + c_2 r_2^n$ para quaisquer valores iniciais $x_1 = a_1$ e $x_2 = a_2$.

Basta entender agora se esta é a única solução da equação para estes valores iniciais.

Seja então $(x_n)_{n \geq 1}$ uma solução qualquer de $x_{n+2} + p x_{n+1} + q x_n = 0$, com $x_1 = a_1$ e $x_2 = a_2$.

Já sabemos que $x_1 = a_1 = c_1 r_1 + c_2 r_2$ e $x_2 = a_2 = c_1 r_1^2 + c_2 r_2^2$, já que c_1 e c_2 foram escolhidos de modo que isso ocorra.

Suponhamos então que $x_k = c_1 r_1^k + c_2 r_2^k$ para todo $k \leq n+1$. Como $x_{n+2} + p x_{n+1} + q x_n = 0$, temos que

$$\begin{aligned} x_{n+2} &= -p(c_1 r_1^{n+1} + c_2 r_2^{n+1}) - q(c_1 r_1^n + c_2 r_2^n) \\ &= -c_1 r_1^n (p r_1 + q) - c_2 r_2^n (p r_2 + q). \end{aligned}$$

Somando e subtraindo $c_1 r_2^{n+2} + c_2 r_2^{n+2}$, encontramos

$$x_{n+2} = -c_1 r_1^n (r_1^2 + p r_1 + q) - c_2 r_2^n (r_2^2 + p r_2 + q) + c_1 r_2^{n+2} + c_2 r_2^{n+2}, \tag{2.8}$$

porém as expressões entre parênteses se anulam, levando a

$$x_{n+2} = c_1 r_2^{n+2} + c_2 r_2^{n+2}. \tag{2.9}$$

Mostramos assim por indução que o polinômio característico da equação linear $x_{n+2} + px_{n+1} + qx_n = 0$ possui duas raízes reais distintas r_1 e r_2 , então para todo $a_1, a_2 \in \mathbb{R}$ existe uma única solução $(x_n)_{n \geq 1}$ tal que $x_1 = a_1$ e $x_2 = a_2$. Além disso,

$$x_n = c_1 r_1^n + c_2 r_2^n,$$

par algum par de constantes $c_1, c_2 \in \mathbb{R}$.

Exemplo 2.8. Determinar as soluções da equação

$$x_{n+2} + 3x_{n+1} - 4x_n = 0. \quad (2.10)$$

A equação característica $r^2 + 3r - 4 = 0$, tem raízes 1 e -4. As soluções da recorrência são as sequência da forma $x_n = c_1 1^n + c_2 (-4)^n$, isto é:

$$x_n = c_1 + c_2 (-4)^n,$$

onde c_1 e c_2 são constantes arbitrárias.

Exemplo 2.9 (A sequência de *Fibonacci*). A sequência de *Fibonacci* é definida por: $f_0 = f_1 = 1$, $f_{n+2} = f_{n+1} + f_n$, para $n \geq 0$.

A equação característica é, portanto, $r^2 - r - 1 = 0$, que tem as raízes

$$r_1 = \frac{1 + \sqrt{5}}{2} \quad \text{e} \quad r_2 = \frac{1 - \sqrt{5}}{2},$$

Logo

$$f_n = c_1 \left(\frac{1 + \sqrt{5}}{2} \right)^n + c_2 \left(\frac{1 - \sqrt{5}}{2} \right)^n,$$

onde os valores de c_1 e c_2 são obtidos usando $f_0 = f_1 = 1$.

Para isso faça

$$\begin{aligned} c_1 + c_2 &= 1, \\ c_1 \frac{1 + \sqrt{5}}{2} + c_2 \frac{1 - \sqrt{5}}{2} &= 1 \end{aligned}$$

que nos dá

$$c_1 = \frac{\sqrt{5} + 1}{2\sqrt{5}} \quad \text{e} \quad c_2 = \frac{\sqrt{5} - 1}{2\sqrt{5}}.$$

Concluimos portanto que

$$\begin{aligned} f_n &= \frac{\sqrt{5}+1}{2\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}-1}{2\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n \\ &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}. \end{aligned}$$

O CASO DE UMA ÚNICA RAÍZ

O caso em que o polinômio característico tem raízes $r_1 = r_2 = r$ nos trás um problema. Apesar de termos uma raiz $r \neq 0$, e portanto toda uma classe de soluções do tipo $x_n = c_1 r^n$, o caso anterior nos mostrou a possibilidade (e possivelmente uma obrigatoriedade) de ao menos mais uma classe de soluções.

Mas que solução seria essa? Já sabemos que ela não tem a forma λ^n , com $\lambda \neq r$, pois tal constante apareceria como raiz do polinômio característico.

Ao mesmo tempo, a forma matricial (2.5) nos aponta na direção de uma potência M^n , que em geral envolve potências, de um modo ou de outro, em todos os seus termos.

Isso indica que a próxima solução não pode estar muito longe.

Antes de seguir vamos olhar um pouco mais de perto a relação entre a raiz r e os coeficientes p, q do polinômio característico. Sabemos (da tão mal falada fórmula de Baskara) que

$$r = \frac{-p \pm \sqrt{p^2 - 4q}}{2}.$$

Como r é a única raiz, temos $p^2 = 4q$, e portanto $r = -\frac{p}{2}$.

Voltando ao problema, vamos agora aplicar uma técnica que usamos em equações de primeira ordem, e buscar uma solução do tipo $x_n = y_n r^n$.

Neste caso, temos

$$y_{n+2} r^{n+2} + p y_{n+1} r^{n+1} + q y_n r^n = 0$$

e portanto

$$y_{n+2} r^2 + p y_{n+1} r + q y_n = 0.$$

Como $r = -p/2$ e $p^2 = 4q$ temos

$$p^2 y_{n+2} - 2p^2 y_{n+1} + p^2 y_n = 0,$$

ou ainda

$$y_{n+2} - 2y_{n+1} + y_n = 0.$$

Encontramos uma nova equação linear homogênea, mas desta vez é uma equação bem específica, que não trás nenhum dos coeficientes originais, tendo surgido apenas do fato do polinômio característico ter apenas uma raiz.

Se tentarmos usar a técnica do polinômio característico para resolver essa equação, cairemos no mesmo problema anterior: temos apenas uma raiz, que neste caso é $r = 1$ (nos dando uma solução constante).

Mas olhando mais atentamente, vemos que é possível decompor a equação de modo mais eficaz. De fato, podemos escrever

$$(y_{n+2} - y_{n+1}) - (y_{n+1} - y_n) = 0,$$

e fazendo $z_n = y_{n+1} - y_n$, obtemos

$$z_{n+1} = z_n,$$

que é uma equação linear de primeira ordem, com solução $z_n = z_1 = c_2 \in \mathbb{R}$.

Assim, temos

$$y_{n+1} = y_n + c_2,$$

de onde concluímos que

$$y_n = c_1 + c_2 n,$$

com $c_1, c_2 \in \mathbb{R}$.

Temos assim que

$$x_n = c_1 \cdot r^n + c_2 \cdot n \cdot r^n$$

é solução da equação buscada.

É interessante observar que fazendo $c_1 = 0$ e $c_2 = 1$ encontramos uma solução da forma $n r^n$, e fazendo $c_1 = 1$ e $c_2 = 0$ recuperamos a solução r^n , indicada pela solução da equação característica. Portanto, assim como no caso de raízes distintas, temos uma solução x_n escrita como combinação de duas soluções individuais distintas.

Cabe então a pergunta: toda solução tem esta forma?

A resposta é positiva, e a prova é análoga ao caso em que $r_1 \neq r_2$.

Primeiro, observamos que dados $a_1, a_2 \in \mathbb{R}$, podemos escolher as constantes c_1 e c_2 de modo que

$$\begin{aligned}c_1 r + c_2 r &= a_1, \\c_1 r^2 + 2c_2 r^2 &= a_2,\end{aligned}$$

(o sistema tem solução única para todo $r \neq 0$). Depois verificamos por indução que se $x_1 = a_1$ e $x_2 = a_2$, então

$$x_n = c_1 r^n + c_2 n r^n,$$

para todo $n \geq 1$.

Exemplo 2.10. Resolver a equação $x_{n+2} - 4x_{n+1} + 4x_n = 0$, com $x_1 = 0$ e $x_2 = 4$.

A equação característica é $r^2 - 4r + 4 = 0$, com raízes $r_1 = r_2 = 2$.

A solução geral da equação é, portanto

$$x_n = c_1 2^n + c_2 n 2^n.$$

Olhando para as condições iniciais, obtemos

$$\begin{cases} 2c_1 + 2c_2 = x_1 = 0 \\ 4c_1 + 8c_2 = x_2 = 4 \end{cases}.$$

Segue então que $c_1 = -1$, $c_2 = 1$ e a solução da equação é

$$x_n = (n - 1)2^n.$$

2.2 PROBABILIDADE

Como comentamos anteriormente, acredita-se que um dos motivos para as ideias de Mendel permanecerem incompreendidas durante mais de três décadas foi a hipótese da aleatoriedade presente na formação dos gametas. A probabilidade não tinha suas bases teóricas bem estabelecidas, e sua aplicação em modelagem de problemas naturais ainda era pouco (ou nada) disseminada.

A probabilidade é a área da matemática que estuda incertezas. Ela forma a base teórica para a Estatística, ciência utilizada nas mais diversas atividades humanas, sendo fundamental

em várias áreas, como Ciências Humanas, Ciências da Saúde, Economia e Finanças, Ecologia e Teoria dos Jogos, dentre outros. Do ponto de vista teórico, atualmente, a Teoria de Probabilidade é utilizada como ferramenta em algumas áreas da Física e, cada vez mais, em áreas da própria Matemática. Por esse motivo, o ensino de Probabilidade no Ensino Médio é importante e atual.

Esse assunto é muito vasto, e aqui trataremos apenas de alguns conceitos básicos e suas aplicações. O leitor interessado em se aprofundar um pouco mais no assunto, pode encontrar um pouco mais de detalhes em [13, 14].

2.2.1 Teoria da probabilidade

A modelagem de incertezas é uma questão antiga na matemática. Pascal e Fermat trocaram diversas cartas discutindo os fundamentos matemáticos de tal problema. Outros nomes famosos, como Bernoulli, Poisson e Gauss, deram contribuições importantes para a área. Mas foi apenas no final do século XIX, com o desenvolvimento da teoria da medida, que chegou-se a uma formulação mais precisa para o que entendemos hoje por probabilidade. Apenas em 1933 Kolmogorov publica o primeiro artigo, estabelecendo de maneira definitiva os conceitos que usamos até hoje.

Abaixo vamos fazer apenas um breve resumo simplificado desta base teórica, nos restringindo apenas aos tipos de espaço usados neste trabalho.

A ideia da probabilidade é simplesmente *modelar a incerteza*. Em outras palavras, representar matematicamente os aspectos incertos de um problema.

Talvez a melhor forma de entender isso é começar com um exemplo.

Considere então o experimento de se lançar um dado. A tentativa de modelar o movimento do dado, desde sua saída da mão do jogador até sua parada na mesa, parece um problema extremamente complexo. Precisaríamos levar em consideração a posição exata do dado no momento da saída, sua velocidade, momento angular, condições de colisão com a superfície da mesa, dentre outras variáveis. Além de altamente complexo, o modelo ainda não atacaria várias das incertezas naturais do problema, como a variabilidade das condições iniciais em cada lançamento.

Uma ideia, já comum à época de Pascal, é atribuir à cada face do dado o valor de $1/6$, representando o que informalmente chamamos de "probabilidade de observar aquela face". Este valor nasce da consideração de que nenhuma das seis faces do dado tem nada de especial em relação às demais, e portanto as condições físicas que favorecem a ocorrência de cada uma delas devem ocorrer com a mesma frequência após "muitas" rolagens.

Assim, matematicamente, o modelo pode ser representado por dois elementos:

- um conjunto $S = \{1, 2, 3, 4, 5, 6\}$;
- uma função $p : S \rightarrow [0, 1]$ dada por $p(k) = 1/6$; $k \in S$.

Com o tempo percebeu-se que era mais prático associarmos os valores de probabilidade não à elementos de S , mas aos seus subconjuntos. Assim conseguimos estudar resultados como, por exemplo, observar um número par, usando a mesma notação matemática.

Com isso passamos a descrever o modelo acima por dois elementos

- um conjunto $S = \{1, 2, 3, 4, 5, 6\}$;
- uma função \mathbb{P} que associa a cada $A \subset S$ o valor

$$\mathbb{P}(A) = \sum_{k \in A} p(k).$$

Assim, se para o resultado de observar um número par, temos $A = \{2, 4, 6\}$ e

$$\mathbb{P}(A) = p(2) + p(4) + p(6) = \frac{1}{3}.$$

2.2.2 Definição de Probabilidade

Para começar a definir o que entenderemos por probabilidade, precisamos começar com o objeto à ser modelado. Chamaremos este objeto genérico de **experimento aleatório**.

O conjunto de todos os resultados possíveis de uma experimento aleatório recebe o nome de **espaço amostral**, que representaremos por S .

Por razões de simplicidade, só vamos considerar aqui o caso onde S é finito ou infinito enumerável.

Os subconjuntos de S serão chamados de **eventos**. Diremos que um evento $A \subset S$ ocorre quando o resultado $x \in S$ do experimento pertence ao evento, ou seja, quando $x \in A$.

Colocados estas definições preliminares, estamos prontos para definir uma probabilidade.

Definição 2.2. Dado um espaço amostral S finito ou infinito enumerável, chamaremos de **probabilidade** em S uma função \mathbb{P} definida nos subconjuntos de S tal que

- $\mathbb{P}(A) \geq 0$ para todo evento $A \subset S$;
- $\mathbb{P}(S) = 1$;
- Se $A, B \subset S$ são tais que $A \cap B = \emptyset$, então $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

No caso que estamos considerando, onde S é finito ou infinito enumerável, a maneira mais comum de montar um modelo é associando à cada elemento $x \in S$ uma probabilidade $\mathbb{P}(\{x\})$. Feito isso, podemos definir $\mathbb{P}(A)$ para qualquer $A \subset S$ como

$$\mathbb{P}(A) = \sum_{x \in A} \mathbb{P}(\{x\}),$$

como fizemos no exemplo dos dados.

Para esclarecer esse conceito, façamos mais um modelo.

Exemplo 2.11 (Acasalamento Aleatório). A coloração em espécie de animais é considerado por muitos biólogos como um fator importante na evolução de espécies de animais, podendo prover vantagens (camuflagem, por exemplo) que auxiliariam na seleção natural da espécie. Tendo este problema em mente, pesquisadores decidiram estudar uma população de gambás, formada por indivíduos amarronzados e negros com manchas brancas. Sabe-se que 60% da população é de machos. Dentre os machos 40% são amarronzados, enquanto dentre as fêmeas essa proporção é 35%.

Nasce um novo filhote, ainda sem pelo, e os pesquisadores querem saber a probabilidade de tal filhote desenvolver o pêlo da cor amarronzada, sabendo que tal coloração se desenvolve se, e só se, quando ambos os progenitores possuem tal cor.

Para responder tal questão precisamos primeiro estabelecer um experimento aleatório, que posteriormente será modelado. Neste caso, podemos pensar que um casal (uma macho e uma fêmea) aleatório se encontrou para acasalar. Com isso denotando por G_m e G_f as populações de gambás machos e fêmeas, respectivamente, podemos descrever o espaço amostral por

$$S = G_m \times G_f = \{(a, b) : a \in G_m \text{ e } b \in G_f\}.$$

Para finalmente chegarmos em um modelo, precisamos fazer algumas hipóteses. Primeiro supomos que não existe influência da cor na escolha de parceiros sexuais nesta espécie de gambá. Segundo que nesta população os indivíduos vivem em locais próximos, sem agrupamentos baseados em cor ou gênero.

Se as hipóteses acima forem aceitas, é razoável pensar em um modelo onde a probabilidade de cada casal $(a, b) \in S$ acasalar é sempre igual. Assim, se a população tem n gambás, com $0,6n$ machos e $0,4n$ fêmeas, temos

$$\mathbb{P}(\{(a, b)\}) = \frac{1}{0,6 \cdot 0,4 \cdot n^2},$$

para todo $(a, b) \in S$.

Agora, com o modelo em mãos, estamos prontos para responder a pergunta. Como o filhote desenvolve cor marrom sempre que os pais possuem a mesma cor, e apenas neste caso, podemos escrever o evento de interesse como

$$M = \{(a, b) : a \in G_{ma} \text{ e } b \in G_{fa}\},$$

onde G_{ma} e G_{fa} são, respectivamente, os conjuntos dos machos amarronzados e das fêmeas amarronzadas.

Assim, como temos $0,4 \cdot 0,6 \cdot n$ indivíduos em G_{ma} e $0,35 \cdot 0,4 \cdot n$ indivíduos em G_{fa} , existem $0,4 \cdot 0,6 \cdot 0,35 \cdot 0,4 \cdot n^2$ indivíduos em M , e portanto

$$\mathbb{P}(M) = \frac{0,4 \cdot 0,6 \cdot 0,35 \cdot 0,4 \cdot n^2}{0,6 \cdot 0,4 \cdot n^2} = 0,4 \cdot 0,35 = 0,14.$$

É interessante observar que, neste modelo, a probabilidade de dois indivíduos amarronzados se acasalarem não dependeu do tamanho da população de machos e fêmeas, e sim das proporção de indivíduos amarronzados em cada grupo.

Observação 2.2.1. *É importante observar que probabilidade, como definido acima, é apenas um objeto matemático com certas propriedades. Tal objeto não surge espontaneamente quando realizamos um dado experimento aleatório. Ao contrário, como em todo modelo, somos nós que devemos escolher que probabilidade melhor representa o problema que estamos enfrentando.*

Como em qualquer problema de modelagem, tal escolha é feita baseada nas hipóteses que levantamos sobre o experimento aleatório, e sua escolha nem sempre é fácil. Se, por exemplo,

estivéssemos modelando a rolagem de um dado viciado, a escolha de probabilidades seria diferente, e seriam necessários mais informações sobre o dado para fazer uma escolha consciente e consistente.

Nos nossos problemas, em especial no próximo capítulo, tentaremos sempre listar as hipóteses que levam ao modelo escolhido.

2.2.3 Propriedades de uma Probabilidade

2.2.4 Probabilidade Condicional

Imagine-se em um jogo de cartas. Não importa exatamente qual, desde que as cartas tenham sido distribuídas no início da partida, definindo a sorte de cada jogador.

Antes das cartas serem distribuídas, você tem alguma noção das probabilidades envolvidas. Você pode tentar calcular as chances de um oponente receber duas das quatro Damas disponíveis, por exemplo. Mas assim que as cartas são distribuídas, tudo isso muda. Se você recebeu uma Dama, a probabilidade do seu oponente ter duas, reduziu um pouco. Se você tem um par, a probabilidade dele estar com o outro é ainda menor, e se você tem um trio de Damas, não tem como ele ter um par.

Estas novas probabilidades, calculadas a partir de uma informação parcial sobre o resultado de um experimento, é conhecida por *probabilidade condicional*, e é uma das ferramentas mais importantes na teoria das probabilidades.

Antes de definir formalmente o que entenderemos por probabilidade condicional, tome dos eventos A e B em um espaço amostral S de um dado experimento aleatório. Imagine que o experimento foi realizado, e que você tenha recebido a informação de que o evento B ocorreu. Ou seja, o resultado do experimento é um elemento de B . Você ainda não sabe qual o resultado, mas quer calcular a probabilidade de A ocorrer, dada a informação de que B ocorreu.

O primeiro passo é descartar todo elemento de A que esteja fora de B . Isso por que sabemos que eles não são candidatos viáveis. Em outras palavras, cada um destes elementos deve ter probabilidade zero, quando condicionamos à ocorrência de B .

Devemos então nos concentrar nos resultados de $A \cap B$.

Além disso, é razoável supor que cada elemento em $A \cap B$ tenha sua probabilidade alterada na mesma proporção. Isso por que, fazer o contrário alteraria o “balanço” entre as probabilidades dos resultados possíveis.

Para entender isso melhor, pense que um dado foi lançado. A princípio, cada resultado tem a mesma probabilidade ($1/6$) de ocorrer. Suponha que você recebe a informação que o número sorteado é par. Isso leva as probabilidades de 1, 3 e 5 para zero, e as demais devem ser alteradas na mesma proporção. Caso contrário teríamos um dos valores mais provável que os demais, e isso adicionaria ao modelo informações ou hipóteses que não temos disponível.

Com isso, é natural pensar que a probabilidade condicional de A dado B , denotado por $\mathbb{P}(A|B)$, é dada por

$$\mathbb{P}(A|B) = \alpha \cdot \mathbb{P}(A \cap B),$$

onde a constante α não depende do evento A , apenas do B .

Agora, basta considerar que $\mathbb{P}(B|B) = 1$, e encontramos a definição abaixo.

Definição 2.3. Dada uma probabilidade \mathbb{P} em um espaço amostral S , e eventos $A, B \subset S$, com $\mathbb{P}(B) > 0$, definimos a probabilidade condicional de A dado B por

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

Exemplo 2.12. Uma população de 200 gambás é dividida em 50 fêmeas de pêlo amarronzado, 60 fêmeas de pelo negro com manchas brancas, 30 machos de pêlo amarronzado e outros 60 negros de manchas brancas.

Um biólogo pede a um estagiário para selecionar uma fêmea do grupo para estudo. Sem saber distinguir o gênero de um gambá, e com vergonha de perguntar, ele pega um gambá aleatório.

Dado que o gamba escolhido tem pêlo amarronzado, qual a probabilidade de ser realmente fêmea.

Para calcular isso, tome os eventos

$$A = \{\text{pêlo amarronzado}\}$$

e

$$F = \{\text{fêmea}\}.$$

Sabemos que

$$\mathbb{P}(A) = \frac{80}{200} = \frac{2}{5},$$

e

$$\mathbb{P}(A \cap F) = \frac{50}{200} = \frac{1}{4}.$$

Assim

$$\mathbb{P}(F|A) = \frac{\mathbb{P}(A \cap F)}{\mathbb{P}(A)} = \frac{1/4}{2/5} = \frac{5}{8}.$$

Assim, se $S = B_1 \cup B_2 \cup \dots \cup B_n$ com $B_i \cap B_j = \emptyset$, sempre que $i \neq j$, então para todo $A \subset S$ podemos escrever

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_n),$$

com $(A \cap B_i) \cap (A \cap B_j) = A \cap B_i \cap B_j = \emptyset$, sempre que $i \neq j$. Segue portanto que

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \dots + \mathbb{P}(A \cap B_n) \\ &= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \dots + \mathbb{P}(A|B_n)\mathbb{P}(B_n). \end{aligned}$$

Exemplo 2.13. Com a intenção de estudar a propagação dos diversos alelos de um gene específico relacionado à coloração dos gambás, um biólogo realiza o seguinte experimento: primeiro ele sorteia aleatoriamente um indivíduo da população de gambás, isolando o locus correspondente no cromossomo, o biólogo observa então apenas um dos alelos, ao acaso. Considerando um gene com alelos A e a , com 45% de indivíduos AA e 35% de indivíduos Aa , queremos saber qual a probabilidade de se observar o alelo a no experimento.

O experimento já está descrito, e tem duas partes. Na primeira observamos o locus, com um par de alelos, em seguida observamos um destes alelos. Com isso podemos representar um resultado possível como um par (a, b) , onde $a \in \{AA, aa, Aa\}$ representa os pares de alelos possíveis, e $b \in \{1, 2\}$ representa o alelo observado. Assim

$$S = \{(AA, 1); (aa, 1); (Aa, 1); (AA, 2); (aa, 2); (Aa, 2)\}.$$

Para escolher \mathbb{P} podemos começar seguindo as proporções conhecidas da população, e supondo que cada indivíduo tem a mesma probabilidade de ser sorteado, determinar que

$$\mathbb{P}(\{(AA, 1); (AA, 2)\}) = 0,45, \quad \mathbb{P}(\{(Aa, 1); (Aa, 2)\}) = 0,35 \quad \text{e} \quad \mathbb{P}(\{(aa, 1); (aa, 2)\}) = 0,2.$$

Para os demais eventos podemos, por exemplo, usar a *probabilidade total*. Com isso, defina primeiro os eventos

$$G_{AA} = \{(AA, 1); (AA, 2)\}, \quad G_{Aa} = \{(Aa, 1); (Aa, 2)\} \quad \text{e} \quad G_{aa} = \{(aa, 1); (aa, 2)\}.$$

Se quisermos calcular agora a probabilidade do evento

$$E_a = \{\text{o alelo observado foi o } a\} = \{(aa, 1); (aa, 2); (Aa, 2)\},$$

podemos fazer

$$\mathbb{P}(E_a) = \mathbb{P}(E_a | G_{AA})\mathbb{P}(G_{AA}) + \mathbb{P}(E_a | G_{aa})\mathbb{P}(G_{aa}) + \mathbb{P}(E_a | G_{Aa})\mathbb{P}(G_{Aa}).$$

Para determinar as condicionais, note que

$$\mathbb{P}(E_a | G_{AA}) = 0 \quad \text{e} \quad \mathbb{P}(E_a | G_{aa}) = 1.$$

Agora, considerando a hipótese de que cada um dos dois alelos no locus isolado tem a mesma probabilidade de ser observado, poder definir que

$$\mathbb{P}(E_a | G_{Aa}) = \frac{\mathbb{P}(\{(Aa, 2)\})}{\mathbb{P}(\{(Aa, 1); (Aa, 2)\})} = \frac{\mathbb{P}(\{(Aa, 2)\})}{\mathbb{P}(\{(Aa, 1)\}) + \mathbb{P}(\{(Aa, 2)\})} = \frac{1}{2}.$$

Assim

$$\mathbb{P}(E_a) = 0 \cdot 0,45 + 1 \cdot 0,2 + \frac{1}{2} \cdot 0,35 = 0,375.$$

MODELOS PARA TRANSMISSÃO DE HERANÇA GENÉTICA

Neste capítulo estudaremos alguns modelos aleatórios simples de transmissão de herança genética em uma população ao longo de gerações, como propostos em [5]. Estudaremos apenas alguns dos modelos ali descritos.

Começaremos com um modelo básico, em uma população monóica, que usaremos para definir o conhecido *equilíbrio de Hardy-Weinberg*. Seguiremos então para uma população dióica, onde estudaremos a transmissão de genótipos ligados à um cromossomo sexual, e sua relação com o equilíbrio de Hardy-Weinberg.

3.1 FREQUÊNCIA GENOTÍPICAS E ALÉLICAS

Antes de começar vamos estabelecer algumas definições e resultados iniciais, necessários no resto do capítulo.

Como foi dito anteriormente, todo modelo aleatório é baseado em um experimento aleatório. Assim, como primeiro passo no estudo, seria necessário estabelecer tal experimento, e descrever ao menos o espaço amostral associado a ele.

No entanto, esta não é exatamente a maneira que começaremos aqui. A verdade é que são muitos os experimentos que podemos descrever, todos adequados ao estudo de transmissão genética. Podemos pensar em um experimento em laboratório, onde indivíduos de uma dada espécie, como uma planta, são cultivados em laboratório, observando todas as hipóteses

desejadas. Por outro lado, podemos pensar em um experimento “natural”, onde os indivíduos são livres, e os dados são colhidos através da escolha de indivíduos na natureza.

Não interessa exatamente o experimento escolhido, os modelos que estudaremos serão sempre baseados em uma hipótese do tipo “sorteio aleatório”. Podemos pensar que quando dois indivíduos se encontram para acasalar, por exemplo, este encontro é feito quase como um sorteio, no sentido que qualquer par compatível teria a mesma probabilidade de se encontrar. Do mesmo modo, quando falamos da probabilidade de se observar um certo genótipo da população, estamos falando de um sorteio aleatório, onde cada indivíduo tem a mesma probabilidade de ser sorteado e observado.

Esta hipótese nos leva a uma relação direta entre as probabilidades dos eventos e a frequência populacional da característica que queremos observar. Neste trabalho estaremos interessados em duas características principais: genótipos (que em populações diplóides são frequentemente representados por um par de alelos em um locus), ou apenas um alelo do locus.

Genótipos representam características genéticas do indivíduos da população, e sua importância é clara. Alelos individuais aparecem durante o processo de reprodução, quando cada genitor contribui com apenas um alelo na formação da prole.

Definição 3.1. Dada uma população onde é possível observar genótipos G_1, \dots, G_k . A frequência genotípica f_{G_1}, \dots, f_{G_k} é dada por

$$f_{G_1, \dots, G_k} = \frac{\text{n}^\circ \text{ de ocorrências dos genótipos } G_1, \dots, G_k}{\text{tamanho da população}} \quad (3.1)$$

Seja l um locus, com alelos A ou a . As frequências alélicas f_A e f_a são dadas por

$$f_A = \frac{\text{n}^\circ \text{ de } A\text{'s na coleção alélica de } l}{\text{tamanho de coleção alélica de } l} \quad (3.2)$$

Exemplo 3.1. Temos dois alelos a e A num locus, para uma população diplóide.

Foram coletadas 15 amostras: AA, Aa, AA .

Nestas condições temos: $f_{AA} = \frac{4}{15}$, $f_{Aa} = \frac{6}{15}$, $f_{aa} = \frac{5}{15}$ e $f_A = \frac{14}{30}$, $f_a = \frac{16}{30}$.

Observe que $f_A + f_a = 1$ e $f_{AA} + f_{Aa} + f_{aa} = 1$.

Como comentamos, estaremos trabalhando com experimentos do tipo “sorteio” de indivíduos. Em nosso modelo, tal “sorteio” vai estar presente, por exemplo, na modelagem do encontro de dois membros de uma espécie para acasalamento. Neste caso podemos pensar em um experimento com duas partes: a primeira consiste da escolha do par de indivíduos que se acasalarão, e uma segunda onde o alelo transmitido à prole por cada genitor é determinado.

Neste contexto, as frequências genotípica e alélica representarão, respectivamente, a probabilidade do indivíduo escolhido ter o genótipo de interesse e a probabilidade de um dado genitor transmitir o alelo em questão.

A interpretação da frequência genotípica parece bastante direta, mas comentaremos um pouco mais sobre a frequência alélica na próxima parte do trabalho.

Por enquanto, exploremos um pouco mais a relação entre tais frequências.

Proposição 3.2. *Considere uma população de espécie diplóide e l um locus que admite dois alelos A e a . Nestas condições, vale que*

$$f_A = f_{AA} + \frac{1}{2}f_{Aa} \quad e \quad f_a = f_{aa} + \frac{1}{2}f_{Aa} \quad (3.3)$$

Demonstração. Agora, se N é o tamanho da população e ela é diplóide, então a coleção de alelos em um locus l tem tamanho $2N$. Assim se $n(g_1g_2)$ com $g_1, g_2 \in \{A, a\}$ representa o total de indivíduos g_1g_2 e, analogamente, $n(g)$ com $g \in \{A, a\}$ representa o total de alelos g na população, temos que

$$n(A) = 2n(AA) + n(Aa),$$

e

$$f_A = \frac{n(A)}{2N} = \frac{2n(AA) + n(Aa)}{2N} = \frac{n(AA)}{N} + \frac{n(Aa)}{2N} = f_{AA} + \frac{1}{2}f_{Aa}.$$

E do mesmo modo

$$f_a = \frac{n(a)}{2N} = \frac{2n(aa) + n(Aa)}{2N} = \frac{n(aa)}{N} + \frac{n(Aa)}{2N} = f_{aa} + \frac{1}{2}f_{Aa}.$$

□

3.2 HIPÓTESES INICIAIS DO MODELOS

Antes de introduzir o primeiro dos modelos que estudaremos, vamos comentar algumas das principais hipóteses presentes nos modelos, e suas consequências.

3.2.1 População Monóica ou Dióica

A primeira distinção importante entre modelos acontece no tipo de reprodução que acontece na população que estamos estudando. Neste trabalho estudaremos dois tipos de populações: as monóicas e as dióicas.

Indivíduos em espécies monóicas possuem órgãos sexuais dos dois sexos, e além da fecundação cruzada, é possível também acontecer a autofecundação. A consequência direta disso é que todo par de indivíduos é compatível, inclusive o “par” formado pelo mesmo indivíduo.

Já as populações dióicas possuem separação de gênero, e a reprodução acontece no encontro de pares de gêneros diferentes.

Com isso, faz-se necessária a separação da população em dois grupos, com as respectivas frequências relativas.

Definição 3.3. Dado um subgrupo S da população, defina por f_{G_1, \dots, G_n}^S a frequência genotípica do genótipo G_1, \dots, G_n dentro do grupo S . Ou seja

$$f_{G_1, \dots, G_n}^S = \frac{\text{total de indivíduos com genótipo } G_1, \dots, G_n \text{ em } S}{\text{total de indivíduos em } S}.$$

Analogamente dado um alelo a em um locus l , f_a^S é a frequência alélica de a em S .

Nas populações monóicas estas separações não se fazem necessárias.

3.2.2 Cromossomos Autossômicos ou Sexuais

Nesta hipótese simplesmente estipula-se se a herança genética de interesse em uma população dióica é autossômica (não ligada aos cromossomos sexuais), ou se é sexual (está ligado à um dos cromossomos sexuais).

Essa é uma hipótese importante, que gera uma mudança significativa no modelo, uma vez que afeta de maneira direta a frequência dos diferentes genótipos nas populações de fêmeas e machos.

3.2.3 Acasalamento Aleatório

A hipótese do *acasalamento aleatório* é uma das responsáveis pelo aspecto temporal dos modelos.

Esta hipótese trata do surgimento de um novo indivíduo pela união aleatória de 2 gametas, e supõe que uma nova prole é gerada em um acasalamento que ocorre aleatoriamente na população, no sentido que qualquer par de indivíduos compatíveis tem a mesma probabilidade de se acasalar. Supõe também que, no processo reprodutivo, cada alelo em um certo locus do progenitor tem a mesma probabilidade de ser transmitido.

Assim, em uma população monóica com N indivíduos, pela hipótese de acasalamento aleatório cada um dos N^2 pares de indivíduos tem probabilidade $1/N^2$ de se encontrar e acasalar.

Em uma população dióica, com $n(f)$ fêmeas e $n(m)$ machos, cada um dos $n(f)n(m)$ pares tem probabilidade $1/(n(f)n(m))$ de se acasalar.

Uma das principais consequências desta hipótese está retratada no seguinte resultado.

Lema 3.4. *Dada uma população dióica, defina p_A^S como a probabilidade de, em um acasalamento aleatório, um dos membros do casal transmitir o alelo A . Nestas condições*

$$p_A^S = f_A^S.$$

Se a população é monóica, p_A representa a probabilidade de um dado indivíduo do par, em um acasalamento aleatório, transmitir o alelo A para a prole, e

$$p_A = f_A.$$

Demonstração. Antes de entrar na prova deste resultado, sugerimos primeiro voltar aos exemplos 2.11 e 2.13, onde exemplificamos exatamente os fenômenos que ocorrem aqui.

Para fixar notação, denote por S_f e S_m , respectivamente, o grupo de fêmeas e machos em uma dada população diplóide. Dado um certo conjunto A , denote por $n(A)$ o total de elementos em A .

Observe agora que o experimento aleatório descrito pela hipótese de acasalamento aleatório tem duas fases: na primeira escolhemos um par de indivíduos de $S_f \times S_m$ para acasalar, em seguida escolhemos, para cada indivíduo, qual alelo (1 ou 2) será transmitido.

Com isso podemos, a princípio, escrever o espaço amostral como

$$S = S_f \times \{1, 2\} \times S_m \times \{1, 2\},$$

onde um elemento $(f, i, m, j) \in S$ é composto pela fêmea f reprodutora e o macho reprodutor m , i o alelo transmitido pela fêmea e j o alelo transmitido pelo macho.

A primeira parte da hipótese diz que cada casal tem a mesma probabilidade de ser escolhido, e a segunda diz que cada alelo tem também a mesma probabilidade de ser escolhido.

Em relação ao nosso espaço amostral, isso quer dizer que

$$\mathbb{P}(\{(f, i, m, j)\}) = \frac{1}{n(S)} = \frac{1}{4n(S_f)n(S_m)}.$$

Considere inicialmente o evento

$$G_{AA}^f = \{\text{a fêmea escolhida possui o par de alelos } AA\},$$

e note que se F_{AA} representa o grupo de fêmeas com genótipo AA então

$$G_{AA}^f = \{(f, i, m, j) : f \in F_{AA}\} = F_{AA} \times \{1, 2\} \times S_m \times \{1, 2\}.$$

Segue que

$$\mathbb{P}(G_{AA}^f) = \frac{n(G_{AA}^f)}{4n(S_f)n(S_m)} = \frac{n(F_{AA})4n(S_m)}{4n(S_f)n(S_m)} = f_{AA}^f,$$

onde $f_{AA}^f := f_{AA}^{S_f}$ é a frequência relativa do genótipo AA entre as fêmeas da população.

Da mesma forma

$$\mathbb{P}(G_{Aa}^f) = f_{Aa}^f, \quad \mathbb{P}(G_{aa}^f) = f_{aa}^f,$$

e

$$\mathbb{P}(G_{AA}^m) = f_{AA}^m, \quad \mathbb{P}(G_{Aa}^m) = f_{Aa}^m, \quad \text{e} \quad \mathbb{P}(G_{aa}^m) = f_{aa}^m.$$

Defina agora

$$E_a^f = \{\text{o alelo transmitido pela fêmea foi o } a\},$$

e note que

$$\mathbb{P}(E_a^f) = p_a^f,$$

que queremos calcular.

Sabemos que

$$\mathbb{P}(E_a^f | G_{AA}^f) = 0, \quad \mathbb{P}(E_a^f | G_{aa}^f) = 1 \quad \text{e} \quad \mathbb{P}(E_a^f | G_{Aa}^f) = \frac{1}{2}.$$

Segue que

$$\begin{aligned} p_a^f &= \mathbb{P}(E_a^f) \\ &= \mathbb{P}(E_a^f | G_{AA}^f) \mathbb{P}(G_{AA}^f) + \mathbb{P}(E_a^f | G_{aa}^f) \mathbb{P}(G_{aa}^f) + \mathbb{P}(E_a^f | G_{Aa}^f) \mathbb{P}(G_{Aa}^f) \\ &= 0 \cdot f_{AA}^f + 1 \cdot f_{aa}^f + \frac{1}{2} \cdot f_{Aa}^f \\ &= f_a^f. \end{aligned}$$

As demais igualdades seguem de modo análogo.

O caso de população monóica é similar, bastando fazer $S_m = S_f = S$. □

Outra forma de calcular $\mathbb{P}(E_a^f)$, por exemplo, seria escrever

$$H_i^f = \{\text{a fêmea transmitiu o alelo } i\}$$

e observar que

$$E_a^f = G_{aa}^f \cup (G_{Aa}^f \cap H_2^f) = (F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\}) \times S_m \times \{1, 2\}.$$

Com isso, temos

$$\mathbb{P}(E_a^f) = \frac{n(F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\})n(S_m)2}{n(S)} = \frac{n(F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\})}{2n(S_f)},$$

e o resultado segue imediatamente.

Mas neste momento queremos apenas focar na igualdade acima, e apontar que nesta mesma notação, teríamos

$$E_a^f \cap E_a^m = (F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\}) \times (M_{aa} \times \{1, 2\} \cup M_{Aa} \times \{2\}),$$

onde M_G representa o grupo dos machos de genótipo $G \in \{AA, aa, Aa\}$.

Assim

$$\begin{aligned} \mathbb{P}(E_a^f \cap E_a^m) &= \frac{n(F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\})n((F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\}))}{n(S)} \\ &= \frac{n(F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\})}{2n(S_f)} \frac{n((F_{aa} \times \{1, 2\} \cup F_{Aa} \times \{2\}))}{2n(S_f)} \\ &= \mathbb{P}(E_a^f) \cdot \mathbb{P}(E_a^m) \\ &= p_a^f \cdot p_a^m \\ &= f_a^f \cdot f_a^m. \end{aligned}$$

Tendo em vista o resultado acima, consideremos os eventos

$$C_{AA} = \{\text{a prole gerada no casamento aleatório tem genótipo } AA\},$$

$$C_{Aa} = \{\text{a prole gerada no casamento aleatório tem genótipo } Aa\}$$

e

$$C_{aa} = \{\text{a prole gerada no casamento aleatório tem genótipo } aa\},$$

E observe que, em uma população dióica,

$$\mathbb{P}(C_{aa}) = \mathbb{P}(E_a^f \cap E_a^m) = f_a^f \cdot f_a^m,$$

e analogamente

$$\mathbb{P}(C_{AA}) = f_A^f \cdot f_A^m \quad \text{e} \quad \mathbb{P}(C_{Aa}) = f_A^f \cdot f_a^m + f_a^f \cdot f_A^m.$$

Em uma população monoica, onde os progenitores são escolhidos de um mesmo grupo de acasalamento, temos:

$$\begin{aligned} \mathbb{P}(C_{AA}) &= (f_A)^2 = (f_{AA} + \frac{1}{2}f_{Aa})^2 \\ \mathbb{P}(C_{aa}) &= (f_a)^2 = (f_{aa} + \frac{1}{2}f_{Aa})^2 \\ \mathbb{P}(C_{Aa}) &= 2f_A f_a = 2(f_{AA} + \frac{1}{2}f_{Aa})(f_{aa} + \frac{1}{2}f_{Aa}). \end{aligned} \tag{3.4}$$

3.2.4 População Infinita

Vimos que, sob a hipótese de acasalamento aleatório, a frequência alélica coincide com a probabilidade de um dos progenitores transmitir o referido alelo.

É interessante observar que a frequência alélica caracteriza uma propriedade da população, enquanto a probabilidade p_A se refere a apenas um indivíduo (ou apenas um casal). A forma como o acasalamento acontece, nos permitiu fazer essa relação, mas quando olhamos para um genótipo, como AA , por exemplo, ao calcular a probabilidade de uma prole possuir este genótipo, vimos em (3.4) que

$$\mathbb{P}(C_{AA}) = (f_A)^2 \neq f_{AA}.$$

Isso é um problema importante para o nosso modelo.

Para resolver essa questão, nos valeremos da repetição. A ideia é que após um número suficientemente grande de acasalamentos, a proporção de indivíduos gerados com dado genótipo deve estar muito próxima da probabilidade de gerar um indivíduo com aquele genótipo.

Para uma analogia simples, pense no experimento de rolar um dado não viciado. Ao rolar uma vez o dado, a probabilidade de observarmos o número 1, por exemplo, é $1/6$. Se repetirmos este experimento milhares de vezes, a proporção de vezes que observamos o valor 1 é de aproximadamente $1/6$, e se fizermos o total de repetições tender ao infinito, este valor seria exatamente $1/6$.

Assim, se a população for “infinita” (no sentido de suficientemente grande), a frequência de proles com um dado genótipo será igual à probabilidade de gerarmos uma prole com aquele genótipo em um acasalamento aleatório.

Ou seja, dado um genótipo $G \in \{AA, Aa, aa\}$,

$$p_G := \mathbb{P}(C_G) = f_G^F, \tag{3.5}$$

onde F é a população de proles gerados por um dado grupo de indivíduos.

Observe que a frequência acima não se refere à toda a população. Aqui estamos separando de alguma maneira um grupo de progenitores das suas proles. Em (3.4), as frequências alélicas e genotípicas que ali aparecem se referem ao grupo de progenitores.

Essa separação ficará mais clara na próxima hipótese.

Observação 3.2.1. *O resultado de convergência de proporções comentado aqui pode ser enunciado e demonstrado de maneira mais formal, mas a dificuldade envolvida vai muito além da proposta deste trabalho, sendo por si só um tema possível de trabalho de conclusão. Vamos nos satisfazer com a noção intuitiva, e aqueles que se interessarem podem procurar pela Lei dos Grande Números, presente na maior parte da literatura sobre o tema, da qual recomendamos [14].*

3.2.5 Gerações não-sobrepostas

Aqui supõe-se uma separação total entre gerações, no sentido que o acasalamento aleatório acontece apenas entre membros de uma dada geração, gerando proles que passam a ser reprodutores na geração seguinte.

Com isso denotaremos por

$$f_G^S(t)$$

a frequência de um genótipo ou alelo G uma geração t .

Esta hipótese, junto com a hipótese de acasalamento aleatório, estabelece a estrutura temporal do modelo, relacionando as frequências da geração $t + 1$ com as da geração t .

3.2.6 Ausência ou Presença de Mutação, Seleção ou Migração

Mutação, seleção e migração são os três fenômenos que enfraquecem a hipótese de acasalamento aleatório, alterando ou a hereditariedade de um certo gene, ou até mesmo as frequências de acasalamento.

A mutação ocorre quando algum gene sofre alterações na sua sequência de DNA, transformando um alelo em outro, e alterando as regras de hereditariedade. Tal hipótese ainda funciona com a hipótese de acasalamento aleatório, mas altera suas consequências, modificando o modelo.

A seleção altera a probabilidade de sobrevivência de um indivíduo antes de se tornar reprodutor, ou sua possibilidade de reprodução. Esta seleção pode ocorrer por razões ambientais,

naturais, fenotípicas ou até mesmo genotípicas. Tudo isso poderia alterar profundamente a hipótese de acasalamento aleatório.

Migração é o fenômeno de entrada ou saída em massa de indivíduos em uma população, alterando drasticamente a variabilidade genética desta. Isso altera a estabilidade necessária para o acasalamento aleatório.

3.3 MODELAGEM

A seguir apresentaremos 3 modelos para transmissão de herança genética. Dois deles, mais simples, modelaram a transmissão para genes autossômicos. Tal hipótese, mesmo para populações dióicas, simplificam muito os processos, que estabilizam rapidamente.

Mas tal simplicidade permite que estudemos o chamado equilíbrio de Hardy-Weinberg, importante na análise do último modelo.

Terminaremos estudando um modelo para genes ligados à cromossomos sexuais. Isso trás uma assimetria ao processo, que o torna mais interessante de ser estudado.

3.3.1 *Modelo básico*

Para este modelo trabalharemos com o seguinte conjunto de hipóteses:

H1 Acasalamento aleatório;

H2 Espécies monóicas;

H3 População infinita;

H4 Gerações não-sobrepostas;

H5 Ausência de mutação, seleção e imigração.

Estamos interessados em estudar as frequências genotípicas e alélicas em um locus l , que admite alelos A ou a .

Assim, $f_{AA}(t)$, $f_{Aa}(t)$ e $f_{aa}(t)$ denotam as frequências genotípicas, e $f_A(t)$ e $f_a(t)$ as frequências alélicas da geração t .

A hipótese de acasalamento aleatório e população infinita nos dá então que

$$f_{AA}(t+1) = (f_A(t))^2 \quad (3.6)$$

$$f_{Aa}(t+1) = 2f_A(t)f_a(t) \quad (3.7)$$

$$f_{aa}(t+1) = (f_a(t))^2 \quad (3.8)$$

Segue daí que

$$f_A(t+1) = f_{AA}(t+1) + \frac{f_{Aa}(t+1)}{2} = f_A^2(t) + \frac{2f_A(t)f_a(t)}{2} = f_A(t)(f_A(t) + f_a(t)), \quad (3.9)$$

mas como $f_A(t) + f_a(t) = 1$, então

$$f_A(t+1) = f_A(t),$$

e

$$f_a(t+1) = f_a(t).$$

Segue que as frequências alélicas estão em equilíbrio neste modelo. Ou seja, $f_a(t)$ e $f_A(t)$ são constantes ao longo de diferentes gerações, com

$$f_A(t) = f_A(0) \quad f_a(t) = f_a(0)$$

Como as frequências alélicas estão relacionadas com as frequências genotípicas, tem-se para todo $t \geq 1$,

$$f_{AA}(t) = f_A^2(t-1) = f_A^2(0), \quad (3.10)$$

$$f_{aa}(t) = f_a^2(t) = f_a^2(0) = (1 - f_A(0))^2 \quad (3.11)$$

e

$$f_{Aa}(t) = 2f_A(t-1)f_a(t-1) = 2f_A(0)f_a(0) = 2f_A(0)(1 - f_A(0)). \quad (3.12)$$

Este tipo de equilíbrio, estabelecido pelas equações (3.10), (3.11) e (3.12) acima, é conhecido como equilíbrio de *Hardy-Weinberg*, que definimos mais geralmente abaixo.

Este equilíbrio, descoberto em 1908 o matemático inglês Godfrey H. Hardy e o médico alemão Wilhem Weinberg, estabelece as relações entre as frequências genotípicas em uma dada população quando não atua sobre ela nenhum fator evolutivo ou ambiental.

Eles determinaram que, sob certas condições, e na ausência de pressões evolutivas, estas frequências deveriam permanecer constantes, e satisfazer certas relações. Isso foi (e ainda é) usado por diversos biólogos para identificar a ação de fatores evolutivos em uma dada espécie.

Definição 3.5. As frequências genotípicas f_{AA} , f_{Aa} e f_{aa} , com $f_{AA} + f_{Aa} + f_{aa} = 1$, estão em equilíbrio de Hardy-Weinberg se existe $p \in \mathbb{R}$, com $0 \leq p \leq 1$, tal modo que $f_{AA} = p^2$, $f_{Aa} = 2p(1 - p)$ e $f_{aa} = (1 - p)^2$.

É importante salientar que o equilíbrio de Hardy-Weinberg no modelo acima, só é garantido para $t \geq 1$. É possível, como veremos mais a frente, que as frequências genotípicas iniciais estejam fora do equilíbrio, mas os acasalamentos redistribuem tais frequências de maneira a entrar em equilíbrio já na geração seguinte.

Mas antes do exemplo, veremos uma maneira mais simples de verificar o equilíbrio de Hardy-Weinberg.

Proposição 3.6. As frequências genotípicas f_{Aa} , f_{AA} e f_{aa} estão no equilíbrio de Hardy-Weinberg se, e somente se, $f_{Aa}^2 = 4f_{AA}f_{aa}$.

Demonstração. Primeiro suponha que as frequências f_{Aa} , f_{AA} e f_{aa} estão no equilíbrio de Hardy-Weinberg, e portanto existe $0 \leq p \leq 1$ tal que

$$f_{AA} = p^2, \quad f_{Aa} = 2p(1 - p) \quad \text{e} \quad f_{aa} = (1 - p)^2.$$

Segue então que

$$f_{Aa}^2 = (2p)^2(1 - p)^2 = 4p^2(1 - p)^2 = 4f_{AA}f_{aa}.$$

Basta mostrar que se $f_{Aa}^2 = 4f_{AA}f_{aa}$, então as frequências f_{Aa} , f_{AA} e f_{aa} estão em equilíbrio de Hardy-Weinberg.

Note que como

$$f_A = f_{AA} + \frac{f_{Aa}}{2} \quad \text{e} \quad f_a = f_{aa} + \frac{f_{Aa}}{2},$$

então

$$f_{AA} = f_A - \frac{f_{Aa}}{2} \quad \text{e} \quad f_{aa} = f_a - \frac{f_{Aa}}{2}.$$

Substituindo em $f_{Aa}^2 = 4f_{AA}f_{aa}$ obtemos

$$\begin{aligned} f_{Aa}^2 &= 4 \left(f_A - \frac{f_{Aa}}{2} \right) \left(f_a - \frac{f_{Aa}}{2} \right) \\ &= 4 \left(f_A f_a - f_A \frac{f_{Aa}}{2} - f_a \frac{f_{Aa}}{2} + \frac{f_{Aa}^2}{4} \right) \\ &= 4f_A f_a - 2f_{Aa}(f_A + f_a) + f_{Aa}^2 \\ &= 4f_A f_a - 2f_{Aa} + f_{Aa}^2 \end{aligned}$$

Segue portanto que

$$2f_{Aa} = 4f_A f_a.$$

Ou ainda

$$f_{Aa} = 2f_A f_a.$$

Observe agora que

$$f_{AA} = f_A - \frac{f_{Aa}}{2} = f_A - f_A f_a = f_A(1 - f_a) = f_A^2,$$

e analogamente

$$f_{aa} = f_a - \frac{f_{Aa}}{2} = f_a - f_A f_a = f_a(1 - f_A) = (1 - f_A)^2,$$

Em resumo, encontramos que

$$f_{AA} = f_A^2,$$

$$f_{aa} = (1 - f_A)^2$$

e

$$f_{Aa} = 2f_A(1 - f_A).$$

Fazendo $p = f_A$, temos $0 \leq p \leq 1$, e

$$f_{AA} = p^2, \quad f_{aa} = (1 - p)^2 \quad \text{e} \quad f_{Aa} = 2p(1 - p).$$

Assim f_{Aa} , f_{AA} e f_{aa} estão em equilíbrio de *Hardy-Weinberg*, como queríamos demonstrar. □

Exemplo 3.2. Suponham válidas as hipóteses do modelo básico, e tome $f_{AA}(0)=0,2$; $f_{Aa}(0) = 0,4$ e $f_{aa}(0) = 0,4$.

Nesta geração inicial a população não está no equilíbrio de *Hardy-Weinberg* uma vez que

$$f_{Aa}^2(0) = (0,4)^2 = 0,16 \neq 0,32 = 4(0,2)(0,4) = 4f_{AA}(0)f_{aa}(0).$$

Para encontrar as frequências genotípicas nas gerações seguintes, primeiro vamos calcular a frequência do alelo A na geração $t = 0$, obtida da seguinte maneira:

$$p = f_A(0) = f_{AA}(0) + \frac{f_{Aa}(0)}{2} = 0,2 + \frac{0,4}{2} = 0,4.$$

Assim, para $t \geq 1$ temos

$$\begin{aligned} f_{AA}(t) &= p^2 = (0,4)^2 = 0,16, \\ f_{Aa}(t) &= 2p(1-p) = 2 \cdot 0,4 \cdot 0,6 = 0,48 \\ f_{aa}(t) &= (1-p)^2 = 0,36. \end{aligned} \tag{3.13}$$

Portanto tem-se o equilíbrio de *Hardy-Weinberg*, com

$$f_{Aa}^2 = (0,48)^2 = 0,2304 = 4 \cdot 0,16 \cdot 0,36 = 4f_{AA}f_{aa}.$$

3.4 MODELO BÁSICO PARA POPULAÇÃO DIÓICA

Para este modelo assumiremos um conjunto de hipóteses muito similar ao do modelo anterior, trocando apenas a hipótese (H2) de população monóica para uma de população dióica. Acrescentaremos também a hipótese de que o gene estudado é autossômico.

Com isso, temos as seguintes hipóteses: Para este modelo trabalharemos com o seguinte conjunto de hipóteses:

H1 Acasalamento aleatório;

H2 Espécies monóicas;

H3 População infinita;

H4 Gerações não-sobrepostas;

H5 Ausência de mutação, seleção e imigração;

H6 Gene autossômico.

A primeira consequência importante de uma população dióica é a separação das frequências em dois grupos, definido pelo sexo. É importante lembrar que indivíduos do sexo feminino carregam um par de cromossomos XX , enquanto indivíduos do sexo masculino carregam o par XY . Com isso, de acordo com a hipótese de acasalamento aleatório, uma prole será do sexo masculino se, e somente se, o pai transmitir o cromossomo Y , o que acontece com probabilidade $1/2$.

Com isso, se denotarmos por $p_G^f(t+1)$ a probabilidade de uma prole com pais da geração t herdar o genótipo G e ser do sexo feminino, e $p_G^m(t+1)$ a mesma probabilidade mas com a prole sendo do sexo masculino, então segue que

$$p_G^m(t+1) = p_G^f(t+1),$$

para todo $t \geq 0$.

A hipótese da população infinita nos dá que

$$f_G^f(t+1) = p_G^f(t+1) \quad \text{e} \quad f_G^m(t+1) = p_G^m(t+1),$$

para todo $t \geq 0$.

Segue portanto que

$$f_G^f(t) = f_G^m(t), \tag{3.14}$$

para $t \geq 1$, e podemos retirar os indicadores f e m , escrevendo apenas $f_G(t)$, para genótipos $G \in \{AA, Aa, aa\}$.

Isso se transmite para quaisquer alelos em um dado locus, e assim as frequências alélicas também são iguais para os distintos sexos. De fato

$$f_a^f(t) = f_{aa}^f(t) + \frac{f_{Aa}^f(t)}{2} = f_{aa}^m(t) + \frac{f_{Aa}^m(t)}{2} = f_a^m(t).$$

Repetindo o mesmo argumento para o alelo A , encontramos que

$$f_g^f(t) = f_g^m(t), \tag{3.15}$$

para todo $t \geq 1$ e qualquer alelo $g \in \{A, a\}$. Da mesma forma que com a frequência genotípica, excluiremos os índices f e m quando desnecessários, escrevendo apenas $f_g(t)$.

Observe que tais relações só são válidas a partir da geração 1, não precisando ser verificadas na geração 0.

Seguindo como no modelo básico, encontramos que

$$f_{AA}(t+1) = f_A^f(t) \cdot f_A^m(t),$$

$$f_{Aa}(t+1) = f_A^f(t) \cdot f_a^m(t) + f_a^f(t) \cdot f_A^m(t)$$

e

$$f_{aa}(t+1) = f_a^f(t) \cdot f_a^m(t),$$

para todo $t \geq 0$.

Considerando apenas $t \geq 1$ e usando (3.14) e (3.15), recuperamos as mesmas equações do modelo básico, e encontramos que

$$f_a^f(t+1) = f_a^m(t+1) = f_a(t+1) = f_a(t) = f_a(1),$$

para todo $t \geq 1$. O mesmo é válido para $f_A(t)$.

Segue assim que

$$\begin{aligned} f_A(1) &= f_{AA}(1) + \frac{f_{Aa}(1)}{2} \\ &= f_A^f(0) \cdot f_A^m(0) + \frac{f_A^f(0) \cdot f_a^m(0)}{2} + \frac{f_a^f(0) \cdot f_A^m(0)}{2} \\ &= \frac{1}{2} \left[f_A^f(0)(f_A^m(0) + f_a^m(0)) + f_A^m(0)(f_A^f(0) + f_a^f(0)) \right] \\ &= \frac{1}{2} \left[f_A^f(0) + f_A^m(0) \right] \\ &= \frac{1}{2} \left[f_{AA}^f(0) + \frac{f_{Aa}^f(0)}{2} + f_{AA}^m(0) + \frac{f_{Aa}^m(0)}{2} \right] \end{aligned}$$

Assim, do mesmo modo que no modelo básico, obtemos para $t \geq 1$

$$f_{AA}(t+1) = f_A(t)^2 = f_A(1)^2 = \left(\frac{1}{2} \left[f_{AA}^f(0) + \frac{f_{Aa}^f(0)}{2} + f_{AA}^m(0) + \frac{f_{Aa}^m(0)}{2} \right] \right)^2,$$

com igualdades análogas para $f_{aa}(t+1)$ e $f_{Aa}(t+1)$.

Determinamos assim que, a partir da geração $t = 2$, as frequências genótípicas estão em equilíbrio de Hardy-Weinberg.

3.5 MODELO PARA GENES ASSOCIADOS A CROMOSSOMOS SEXUAIS

Genes associados à cromossomos sexuais trazem uma modificação essencial ao modelo.

Considere, por exemplo, seres humanos e suponha que o gene está ligado ao cromossomo Y. Neste caso, independente do alelo considerado, indivíduos do sexo feminino (XY) não carregam o gene, enquanto indivíduos do sexo masculino (XY) carregam o mesmo alelo do pai. Neste caso as frequências para o sexo feminino são sempre nulas, enquanto para o sexo masculino as frequências $f_{AA}^m(t)$, $f_{Aa}^m(t)$ e $f_a^m(t)$ não fazem sentido e

$$f_A^m(t+1) = f_A^m(t) = f_A^m(0) \quad \text{e} \quad f_a^f(t+1) = f_a^f(t) = f_a^f(0),$$

para todo $t \geq 0$.

Para genes ligados ao cromossomo X o modelo fica mais interessante. Neste caso, as frequências $f_{AA}^m(t)$, $f_{Aa}^m(t)$ e $f_a^m(t)$ continuam não fazendo sentido, uma vez que um indivíduo do sexo masculino só carrega o alelo transmitido pela mãe. E pela mesma razão temos

$$f_A^m(t+1) = f_A^f(t) \quad \text{e} \quad f_a^m(t+1) = f_a^f(t), \quad (3.16)$$

para todo $t \geq 0$.

Temos ainda que

$$\begin{aligned} f_{AA}^f(t+1) &= f_A^f(t) \cdot f_A^m(t), \\ f_{Aa}^f(t+1) &= f_A^f(t) \cdot f_a^m(t) + f_a^f(t) \cdot f_A^m(t) \end{aligned}$$

e

$$f_{aa}^f(t+1) = f_a^f(t) \cdot f_a^m(t),$$

para todo $t \geq 0$.

Segue daí que

$$\begin{aligned} f_A^f(t+1) &= f_{AA}^f(t+1) + \frac{f_{Aa}^f(t+1)}{2} \\ &= f_A^f(t) \cdot f_A^m(t) + \frac{f_A^f(t) \cdot f_a^m(t)}{2} + \frac{f_a^f(t) \cdot f_A^m(t)}{2} \\ &= \frac{1}{2} \left[f_A^f(t) \cdot (f_A^m(t) + f_a^m(t)) + f_A^m(t) \cdot (f_A^f(t) + f_a^f(t)) \right] \\ &= \frac{1}{2} \cdot f_A^f(t) + \frac{1}{2} \cdot f_A^m(t). \end{aligned}$$

Concluimos que

$$f_A^f(t+1) = \frac{1}{2} \cdot f_A^f(t) + \frac{1}{2} \cdot f_A^m(t), \quad (3.17)$$

e de (3.16)

$$f_A^f(t+1) = \frac{1}{2} \cdot f_A^f(t) + \frac{1}{2} \cdot f_A^f(t-1),$$

para $t \geq 1$.

Esta é uma equação linear de ordem 2, com equação característica dada por

$$r^2 - \frac{1}{2}r - \frac{1}{2} = 0,$$

que possui raízes $r_1 = 1$ e $r_2 = -\frac{1}{2}$.

Segue daí que existem constantes $c_1, c_2 \in \mathbb{R}$ tais que

$$f_A^f(t) = c_1 + c_2 \left(-\frac{1}{2}\right)^t.$$

Para calcular c_1 e c_2 fazemos

$$\begin{cases} c_1 + c_2 & = f_A^f(0) \\ c_1 - \frac{1}{2}c_2 & = f_A^f(1), \end{cases}$$

o que nos dá, por (3.17), que

$$c_1 = \frac{1}{3} \left(2f_A^f(1) + f_A^f(0) \right) = \frac{1}{3} \left(2f_A^f(0) + f_A^m(0) \right),$$

e

$$c_2 = f_A^f(0) - \frac{1}{3} \left(2f_A^f(0) + f_A^m(0) \right) = \frac{1}{3} \left(f_A^f(0) - f_A^m(0) \right).$$

E portanto

$$f_A^f(t) = \frac{1}{3} \left(2f_A^f(0) + f_A^m(0) \right) + \frac{1}{3} \left(f_A^f(0) - f_A^m(0) \right) \left(-\frac{1}{2}\right)^t, \quad (3.18)$$

para $t \geq 0$.

Para calcular $f_{AA}^f(t)$, $f_{Aa}^f(t)$ e $f_{aa}^f(t)$ fazemos

$$f_{AA}^f(t) = f_A^f(t-1) \cdot f_A^m(t-1) = f_A^f(t-1) \cdot f_A^f(t-2), \quad (3.19)$$

e, analogamente

$$f_{Aa}^f(t) = f_A^f(t-1) \cdot f_a^f(t-2) + f_a^f(t-1) \cdot f_A^f(t-2), \quad (3.20)$$

e

$$f_{aa}(t) = f_a^f(t-1) \cdot f_a^m(t-1) = f_a^f(t-1) \cdot f_a^f(t-2), \quad (3.21)$$

para todo $t \geq 2$.

Não é difícil, usando a Proposição 3.6, ver que as frequências acima não estão no equilíbrio de Hardy-Weinberg. Mas olhando com mais cuidado a equação (3.18), fazendo t ir ao infinito, encontramos

$$f_A^f(\infty) := \lim_{t \rightarrow \infty} f_A^f(t) = \frac{1}{3} \left(2f_A^f(0) + f_A^m(0) \right).$$

E de (3.19), (3.20) e (3.21), segue

$$f_{AA}^f(\infty) := \lim_{t \rightarrow \infty} f_{AA}^f(t) = \left(f_A^f(\infty) \right)^2,$$

$$f_{Aa}^f(\infty) := \lim_{t \rightarrow \infty} f_{Aa}^f(t) = 2f_A^f(\infty) \cdot f_a^f(\infty) = 2f_A^f(\infty) \cdot \left(1 - f_A^f(\infty) \right),$$

e

$$f_{aa}^f(\infty) := \lim_{t \rightarrow \infty} f_{aa}^f(t) = \left(f_a^f(\infty) \right)^2 = \left(1 - f_A^f(\infty) \right)^2,$$

que estão em equilíbrio de Hardy-Weinberg com $p = f_A^f(\infty)$.

Vimos assim que o fato do gene de interesse estar ligado à um cromossomo sexual (no caso, o feminino), não permite que a população entre em equilíbrio imediatamente, como nos demais modelos. Mas ainda assim, com o passar do tempo as frequências se aproximam do equilíbrio, e com velocidade bastante alta.

BIBLIOGRAFIA

- [1] *Minha Vida - Síndrome de Turner*, <http://www.minhavidacom.br/saude/temas/sindrome-de-turner>, acesso em 05-11-2019.
- [2] Mariana Araguaia, *Mundo Educação - Genética*, <https://mundoeducacao.bol.uol.com.br/biologia/genetica.htm>, acesso em 05-11-2019.
- [3] Isabela Canhas, *InfoEscola - Cromossomos*, <https://www.infoescola.com/biologia/cromossomos>, acesso em 05-11-2019.
- [4] Paulo Cezar Pinto Carvalho e Augusto Cezar de Oliveira Morgado, *Matemática Discreta*, Coleção Profmat, SBM, 2015.
- [5] Vinícius Freitas de Oliveira e Suzete Maria Silva Afonso, *Modelos Matemáticos na genética de População*, Revista Eletrônica Paulista de Matemática **13** (2018), 70–89.
- [6] Fabiana Santos Gonçalves, *InfoEscola - Síndrome de Klinefelter*, <https://www.infoescola.com/doencas-geneticas/sindrome-de-klinefelter>, acesso em 05-11-2019.
- [7] Luiz Carlos Uchoa Junqueira e José Carneiro, *Biologia Celular e Molecular*, 9^a ed., Guanabara Koogan, 2007.
- [8] Leonardo Leite, *Genética Clínica - Síndrome de Down*, <http://www.ghente.org/ciencia/genetica/down.htm>, acesso em 05-11-2019.
- [9] _____, *Genética Clínica - Síndrome de Edwards ou Trissomia do 18*, <http://www.ghente.org/ciencia/genetica/trissomia18.htm>, acesso em 05-11-2019.
- [10] _____, *Genética Clínica - Síndrome de Patau ou Trissomia do 13*, <http://www.ghente.org/ciencia/genetica/trissomia13.htm>, acesso em 05-11-2019.
- [11] Lana Magalhães, *Toda Matéria - Herança Ligada ao Sexo*, <https://www.todamateria.com.br/heranca-ligada-ao-sexo/>, acesso em 05-11-2019.

- [12] Débora Carvalho Meldau, *InfoEscola - Síndrome de Edwards*, <https://www.infoescola.com/doencas/sindrome-de-edwards>, acesso em 05-11-2019.
- [13] Sheldon M. Ross, *Introduction to Probability Models*, Probability and Statistics, Academic Press, 2007, ISBN 9780123736352.
- [14] _____, *Probabilidade: Um Curso Moderno com Aplicações*, 8ª ed., Bookman, Porto Alegre, 2010, ISBN 9788577806881.