

ANÁLISE DE DADOS BAYESIANA

SARAH G. A. BARBOSA

2024

SUMÁRIO

I Fundamentos da Inferência Bayesiana

1

1

PROBABILIDADE E INFERÊNCIA	2
1.1 Os Três Passos da Análise de Dados Bayesiana	2
1.2 Notação Geral para Inferência Estatística	3
1.2.1 Parâmetros, Dados e Previsões	3
1.2.2 Unidades Observacionais e Variáveis	4
1.2.3 Permutabilidade	4
1.2.4 Variáveis Explicativas	4
1.2.5 Modelagem Hierárquica	4
1.3 Inferência Bayesiana	5
1.3.1 Regra de Bayes	5
1.3.2 Previsão	6
1.3.3 Verossimilhança	6
1.3.4 Razão de Verossimilhança e Odds Ratio	6
1.4 Exemplos Discretos	7
1.4.1 Inferência sobre o Status Genético	7
1.4.2 Correção Ortográfica	8
RESOLUÇÃO DOS PROBLEMAS	9
Capítulo 1	9
REFERÊNCIAS BIBLIOGRÁFICAS	10

PARTE I: FUNDAMENTOS DA INFERÊNCIA BAYESIANA

1

PROBABILIDADE E INFERÊNCIA

1.1 Os Três Passos da Análise de Dados Bayesiana

Este livro trata de métodos práticos para realizar inferências a partir de dados, utilizando modelos de probabilidade que nos ajudam a entender tanto as quantidades que observamos quanto aquelas sobre as quais queremos aprender mais. A análise Bayesiana é conhecida por aplicar a teoria da probabilidade de forma explícita para quantificar incertezas nas inferências feitas com base nos dados estatísticos.

Podemos dividir o processo de análise de dados Bayesiana em três etapas principais:

1. **Criação de um modelo de probabilidade completo** — Este modelo é uma distribuição conjunta de probabilidade que abrange todas as variáveis, tanto observáveis quanto não observáveis, relacionadas ao estudo. É fundamental que esse modelo esteja alinhado com o conhecimento existente sobre o problema científico em questão e com os métodos usados para coletar os dados.
2. **Condicionamento nos dados observados** — Depois de coletar os dados, movemos para calcular e interpretar a distribuição a posteriori, que é a probabilidade condicional das variáveis não observadas, baseada nos dados que observamos.
3. **Avaliação do modelo e das conclusões derivadas** — Esta etapa envolve verificar o quão bem o modelo se ajusta aos dados coletados, avaliar se as conclusões tiradas são plausíveis e entender a sensibilidade dos resultados às hipóteses feitas na primeira etapa. Se necessário, o modelo pode ser ajustado ou expandido, e as etapas anteriores são repetidas.

Ao longo dos últimos quarenta anos, grandes progressos foram feitos nessas áreas, e muitos deles são discutidos e aplicados em exemplos práticos ao longo do livro. Nosso tratamento cobre todas as três etapas, com o segundo passo focando em métodos computacionais e o terceiro balanceando técnica e julgamento, guiado pelo contexto específico do problema.

A primeira etapa, a construção do modelo, ainda representa um grande desafio: De onde vêm nossos modelos? Como procedemos para criar especificações de probabilidade adequadas? Oferecemos orientações sobre essas questões e enfatizamos a importância da terceira etapa na avaliação retrospectiva do ajuste dos modelos. Avanços no cálculo de distribuições de probabilidade condicionais facilitam o manejo da dependência das conclusões em distribuições a priori subjetivas.

Uma das razões principais para adotar o pensamento Bayesiano é que ele torna as conclusões estatísticas mais intuitivas. Por exemplo, um intervalo de probabilidade Bayesiano para uma quantidade desconhecida pode ser visto diretamente como tendo uma alta probabilidade de incluir essa quantidade. Isso contrasta com um intervalo de confiança frequentista, que é geralmente interpretado com base na ideia de que, se repetíssemos o mesmo experimento muitas vezes, o intervalo incluiria a quantidade desconhecida na maioria das repetições. Recentemente, na estatística aplicada, tem-se dado mais ênfase à estimação por intervalos do que aos testes de hipóteses, reforçando a preferência pelo enfoque Bayesiano, pois é comum que os usuários interpretem os intervalos de confiança de uma maneira intuitiva e Bayesiana. Um dos objetivos

deste livro é mostrar até que ponto essas interpretações Bayesianas de procedimentos estatísticos comuns são justificáveis.

Ao invés de debater sobre os fundamentos teóricos da estatística, focamos nas vantagens práticas da abordagem Bayesiana. Sua flexibilidade e abrangência são essenciais para lidar com problemas complexos. O ponto central da inferência Bayesiana é a capacidade de quantificar diretamente a incerteza, o que, em teoria, permite ajustar modelos complexos, com muitos parâmetros e várias camadas de especificações probabilísticas. Na prática, os desafios incluem configurar e computar esses modelos extensos, e dedicamos uma grande parte deste livro a explorar técnicas recentes e em desenvolvimento que ajudam a superar esses desafios. A abordagem Bayesiana oferece uma maneira conceitualmente simples de gerenciar múltiplos parâmetros, o que discutiremos em detalhes a partir do Capítulo 3.

1.2 Notação Geral para Inferência Estatística

A inferência estatística foca em extrair conclusões sobre quantidades que não podemos observar diretamente, a partir de dados numéricos coletados. Um exemplo prático é o ensaio clínico de um novo medicamento para tratamento de câncer, onde o objetivo é comparar a probabilidade de sobrevivência de cinco anos entre pacientes que receberam o novo medicamento e aqueles que seguiram o tratamento padrão. Essas probabilidades de sobrevivência são estimativas que se referem a uma grande **população** de pacientes e não é viável, nem ético, realizar o experimento em toda essa população. Portanto, as conclusões sobre as verdadeiras probabilidades de sobrevivência e suas diferenças devem ser inferidas de uma **amostra** representativa de pacientes. Além disso, mesmo que fosse possível tratar toda a população com um ou outro tratamento, nunca seria viável expor o mesmo indivíduo a ambos os tratamentos. Assim, a inferência estatística é essencial para fazer comparações **causais** – isto é, comparar o resultado observado em cada paciente com o resultado hipotético que teriam se recebessem o outro tratamento.

Existem dois tipos principais de **estimativas** nas quais a inferência estatística é baseada: primeiro, as quantidades potencialmente observáveis, como futuras observações de um processo ou o resultado sob um tratamento não recebido, como no exemplo do ensaio clínico; e segundo, quantidades que não são diretamente observáveis, como os parâmetros que definem o processo teórico que produz os dados observados (por exemplo, coeficientes de regressão). Essa distinção nem sempre é clara, mas é geralmente útil para compreender como um modelo estatístico é aplicado a problemas reais.

1.2.1 Parâmetros, Dados e Previsões

Na notação que adotamos para inferência estatística, usamos símbolos específicos para representar diferentes tipos de quantidades relacionadas aos dados e aos modelos:

- θ representa **parâmetros** populacionais não observáveis ou quantidades vetoriais de interesse. Por exemplo, no caso do ensaio clínico, θ pode representar as probabilidades de sobrevivência sob cada tratamento para indivíduos aleatoriamente selecionados da população.
- y é usado para denotar os dados observados. Nosso exemplo do ensaio clínico utilizaria y para indicar o número de sobreviventes e mortes em cada grupo de tratamento.
- \tilde{y} representa quantidades desconhecidas, mas potencialmente observáveis. Isso inclui, por exemplo, os resultados hipotéticos de pacientes que recebem o tratamento alternativo ou os resultados futuros de novos pacientes que possam ser semelhantes aos já incluídos no estudo.

Esses símbolos, geralmente, representam quantidades multivariadas, o que significa que podem conter várias medidas ou valores simultâneos. Então, usamos letras gregas para parâmetros, letras romanas minúsculas para dados escalares, vetores observados ou observáveis e letras romanas maiúsculas para representar matrizes ou observáveis. Ao usar a notação matricial, consideramos vetores como vetores coluna por padrão. Por exemplo, se u é um vetor de n componentes, então $u^T u$ (o produto da transposta de u por si mesmo) resulta em um escalar, enquanto $u u^T$ forma uma matriz de dimensão $n \times n$.

1.2.2 Unidades Observacionais e Variáveis

Em muitos estudos estatísticos, os dados são coletados de um conjunto de n objetos ou **unidades**, cada um contribuindo com informações que são organizadas em um vetor $y = (y_1, \dots, y_n)$. No exemplo do ensaio clínico, cada y_i pode ser 1 se o paciente i estiver vivo após cinco anos, ou 0 se o paciente não estiver. Quando múltiplas variáveis são medidas em cada unidade, y_i se torna um vetor, e o conjunto completo de dados y é então considerado uma matriz (comumente com n linhas). Essas variáveis y são chamadas de “resultados” e são tratadas como “aleatórias”, o que significa que, ao fazer inferências, consideramos a possibilidade de que os valores observados poderiam ter sido diferentes, devido ao processo de amostragem e à variabilidade natural da população.

1.2.3 Permutabilidade

A base inicial de uma análise estatística frequentemente parte da suposição (geralmente implícita) de que os valores y_i para cada unidade i podem ser **permutáveis**. Isso significa que expressamos nossa incerteza através de uma densidade de probabilidade conjunta $p(y_1, \dots, y_n)$ que permanece inalterada se os índices dos indivíduos forem permutados. Um modelo não permutável seria apropriado se a informação relevante para os resultados estivesse de alguma forma vinculada aos índices das unidades, em vez de ser explicada por variáveis específicas (como veremos abaixo). A ideia de permutabilidade é fundamental na teoria estatística e será um conceito recorrente ao longo deste livro.

Normalmente, modelamos dados de uma distribuição permutável assumindo que eles são **independentes e identicamente distribuídos** (IID), dados um vetor de parâmetros desconhecidos θ com distribuição $p(\theta)$. No exemplo do ensaio clínico, poderíamos modelar os resultados y_i como IID, dado θ , que representa a probabilidade desconhecida de sobrevivência.

1.2.4 Variáveis Explicativas

Em estudos estatísticos, frequentemente encontramos variáveis em cada unidade que não são modeladas como aleatórias. No exemplo do ensaio clínico, essas variáveis poderiam ser a idade e o estado de saúde anterior dos pacientes. Essas variáveis são denominadas **variáveis explicativas**, e as representamos com x . Usamos X para indicar o conjunto completo de variáveis explicativas para todas as n unidades; se existem k variáveis explicativas, então X é uma matriz com n linhas e k colunas. Considerando X como não aleatório, a noção de permutabilidade pode ser ampliada para exigir que a distribuição dos valores de $(x, y)_i$ seja inalterada por permutações arbitrárias dos índices. É sempre apropriado assumir um modelo permutável após incorporar informações relevantes suficientes em X de modo que os índices possam ser considerados como atribuídos aleatoriamente.

Devido à suposição de permutabilidade, podemos dizer que a distribuição de y dado x é consistente em todas as unidades de estudo no sentido de que se duas unidades compartilham o mesmo x , então elas têm distribuições idênticas de y . Qualquer uma das variáveis explicativas x pode ser tratada como uma variável dependente y se escolhermos modelá-la dessa forma. O papel das variáveis explicativas (também conhecidas como **preditores**) é discutido em detalhes no Capítulo 8, no contexto de análises de pesquisas, experimentos e estudos observacionais, e nas seções posteriores deste livro, no contexto de modelos de regressão.

1.2.5 Modelagem Hierárquica

Nos Capítulos 5 e subsequentes, focamos em **modelos hierárquicos**, que são úteis quando as informações estão disponíveis em vários níveis diferentes de unidades observacionais. Em um modelo hierárquico, podemos aplicar a ideia de permutabilidade em cada nível de unidade. Por exemplo, imagine que dois tratamentos médicos são aplicados em experimentos randomizados distintos em pacientes de várias cidades diferentes. Sem outras informações, seria razoável considerar os pacientes de cada cidade como permutáveis, bem como os resultados das diferentes cidades. Na prática, seria sensato incluir no nível da cidade, como variáveis explicativas, quaisquer informações relevantes que possuímos sobre cada cidade, além das

variáveis explicativas já mencionadas no nível individual. Assim, as distribuições condicionais, dadas essas variáveis explicativas, seriam consideradas permutáveis.

1.3 Inferência Bayesiana

A inferência Bayesiana fundamenta-se em declarações de **probabilidade** para tirar conclusões sobre um parâmetro θ ou sobre dados não observados \tilde{y} . Essas declarações são sempre condicionais aos dados observados y e, na nossa notação, são expressas como $p(\theta|y)$ ou $p(\tilde{y}|y)$. Além disso, há um condicionamento implícito aos valores conhecidos das variáveis explicativas x . Esta abordagem difere significativamente da inferência estatística convencional encontrada em muitos livros-texto, que geralmente se baseia numa avaliação retrospectiva do método usado para estimar θ (ou \tilde{y}) considerando uma distribuição de possíveis valores de y condicionados ao verdadeiro valor desconhecido de θ . Apesar dessa diferença, muitas vezes, análises simples acabam por apresentar conclusões semelhantes em ambas abordagens. Contudo, as análises feitas com métodos Bayesianos podem ser mais facilmente adaptadas para lidar com problemas complexos. Nesta seção, introduzimos a matemática básica e a notação utilizada na inferência Bayesiana, que será ilustrada com um exemplo prático na próxima seção, focado na genética.

Alguns comentários sobre a notação de probabilidade são necessárias aqui. Primeiramente, $p(\cdot|\cdot)$ indica uma **densidade de probabilidade condicional**, cujos argumentos são definidos pelo contexto, enquanto $p(\cdot)$ representa uma **distribuição marginal**. Os termos “distribuição” e “densidade” são usados de forma intercambiável, aplicáveis tanto a funções de densidade contínuas quanto a funções de massa de probabilidade discretas. Nas equações, diferentes distribuições são indicadas pela mesma notação $p(\cdot)$. Embora possa parecer um abuso da notação matemática padrão, esse método proporciona concisão e alinha-se à prática comum de usar $p(\cdot)$ para a probabilidade de qualquer evento discreto, onde o **espaço amostral** é omitido na notação. Dependendo do contexto, para evitar confusões, podemos alternativamente usar a notação $\Pr(\cdot)$ para probabilidade de um evento. Quando referimos uma distribuição padrão, usamos a notação baseada no nome da distribuição; por exemplo, se θ segue uma distribuição normal com média μ e variância σ^2 , escreveríamos $\theta \sim N(\mu, \sigma^2)$ ou $p(\theta) = N(\theta|\mu, \sigma^2)$ ou, para maior clareza, $p(\theta|\mu, \sigma^2) = N(\theta|\mu, \sigma^2)$. Utilizamos notações como $N(\mu, \sigma^2)$ para variáveis aleatórias e $N(\theta|\mu, \sigma^2)$ para funções de densidade.

1.3.1 Regra de Bayes

Para expressar probabilidades sobre a variável θ com base na observação y , é essencial iniciar com um modelo que estabeleça uma **distribuição de probabilidade conjunta** para θ e y . Esta distribuição é geralmente formulada como o produto de duas distribuições individuais, frequentemente chamadas de **distribuição a priori** $p(\theta)$ e a **distribuição amostral** $p(y|\theta)$, respectivamente:

$$p(\theta, y) = p(\theta)p(y|\theta).$$

Ao condicionarmos esses dados ao valor observado de y , e aplicarmos a **regra de Bayes**, obtemos a **distribuição a posteriori**:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}, \quad (1.1)$$

onde,

$$p(y) = \sum_{\theta} p(\theta)p(y|\theta),$$

representa a soma sobre todos os possíveis valores de θ ou,

$$p(y) = \int p(\theta)p(y|\theta) d\theta,$$

se θ for contínuo. Uma versão equivalente da **Equação 1.1**, que ignora o fator $p(y)$ (uma vez que não depende de θ e pode ser tratado como uma constante com y fixo), resulta em uma expressão para a **densidade a posteriori não-normalizada**:

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (1.2)$$

O segundo termo dessa expressão, $p(y|\theta)$, é visto como uma função de θ , e não de y . Essas equações encapsulam a essência técnica da inferência Bayesiana: a principal tarefa em qualquer aplicação é desenvolver adequadamente o modelo $p(\theta, y)$ e executar os cálculos necessários para resumir $p(\theta|y)$ de forma eficaz e apropriada.

1.3.2 Previsão

Ao realizar inferências sobre uma variável desconhecida, utilizamos um raciocínio parecido ao anterior. Antes de considerar os dados y , a distribuição do observável desconhecido y é expressa como:

$$p(y) = \int p(y, \theta) d\theta = \int p(\theta) p(y|\theta) d\theta. \quad (1.3)$$

Esta é conhecida como a **distribuição marginal** de y , mas uma nomenclatura mais descritiva seria a **distribuição preditiva a priori**: “a priori” porque não depende de uma observação prévia, e “preditiva” porque se refere à distribuição de uma variável que ainda será observada.

Após os dados y serem observados, podemos fazer previsões sobre um novo observável desconhecido, \tilde{y} , originário do mesmo processo. Por exemplo, se $y = (y_1, \dots, y_n)$ representa um conjunto de medidas de peso de um objeto n vezes em uma balança, onde $\theta = (\mu, \sigma^2)$ representa o peso verdadeiro desconhecido do objeto e a variância das medições, \tilde{y} poderia ser o peso do objeto numa futura pesagem ainda não realizada. A distribuição de \tilde{y} é chamada de **distribuição preditiva a posteriori**, “a posteriori” porque é condicional aos dados y observados, e “preditiva” porque se refere a uma previsão para um \tilde{y} observável:

$$p(\tilde{y}|y) = \int p(\tilde{y}, \theta|y) d\theta = \int p(\tilde{y}|\theta, y) p(\theta|y) d\theta = \int p(\tilde{y}|\theta) p(\theta|y) d\theta. \quad (1.4)$$

A segunda e a terceira expressão exibem a distribuição preditiva a posteriori como uma média das previsões condicionais baseadas na distribuição posterior de θ . O último passo se dá pela suposta independência condicional entre y e \tilde{y} quando θ é conhecido.

1.3.3 Verossimilhança

Quando aplicamos a regra de Bayes usando um modelo probabilístico escolhido, os dados y influenciam a inferência a posteriori (**Equação 1.2**), apenas através de $p(y|\theta)$. Esta função, ao ser vista como dependente de θ com y fixo, é conhecida como **função de verossimilhança**. Portanto, a inferência Bayesiana adere ao que é frequentemente denominado **princípio da verossimilhança**.

Este princípio é coerente dentro do contexto do modelo ou da família de modelos que escolhemos para uma análise específica. Entretanto, na prática, é raro ter certeza absoluta de que o modelo selecionado é completamente correto. Discutiremos no Capítulo 6 como as distribuições amostrais (considerando repetições dos nossos dados) podem ser fundamentais para verificar as suposições feitas no modelo. De fato, nossa visão sobre um estatístico Bayesiano é a de alguém flexível, disposto a aplicar a regra de Bayes sob uma variedade de modelos possíveis.

1.3.4 Razão de Verossimilhança e Odds Ratio

A razão entre as densidades posteriores $p(\theta|y)$, avaliadas em dois pontos θ_1 e θ_2 dentro de um modelo específico, é chamada de **razão de chances** (em inglês, **odds ratio**, doravante **odds**) a posteriori para θ_1 em comparação a θ_2 . A forma mais comum de aplicar este conceito ocorre quando se considera parâmetros discretos, sendo θ_2 o complemento de θ_1 . As odds representam uma forma alternativa de expressar probabilidades e possuem uma característica atrativa: a regra de Bayes se simplifica bastante quando expressa em termos de odds:

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)/p(y)}{p(\theta_2)p(y|\theta_2)/p(y)} = \frac{p(\theta_1)}{p(\theta_2)} \frac{p(y|\theta_1)}{p(y|\theta_2)} \quad (1.5)$$

Em outras palavras, as odds a posteriori são as odds a priori multiplicados pela **razão de verossimilhança**,

$$\frac{p(y|\theta_1)}{p(y|\theta_2)}.$$

1.4 Exemplos Discretos

A seguir, demonstramos o teorema de Bayes com dois exemplos nos quais o objetivo imediato é a inferência sobre uma quantidade discreta específica, em vez da estimativa de um parâmetro que descreve uma população inteira. Esses exemplos discretos nos permitem ver diretamente as probabilidades a priori, de verossimilhança e a posteriori.

1.4.1 Inferência sobre o Status Genético

Humanos do sexo masculino têm um cromossomo X e um Y, enquanto mulheres possuem dois cromossomos X. A hemofilia é uma doença ligada ao cromossomo X e se manifesta de forma recessiva. Isso significa que um homem com o gene defeituoso no seu cromossomo X será afetado pela doença. Mulheres, por outro lado, precisam herdar o gene defeituoso de ambos os cromossomos X para manifestar a doença, o que é raro, já que a frequência do gene é baixa.

Distribuição a Priori. No caso de uma mulher cujo irmão é afetado pela hemofilia, sabemos que sua mãe é portadora do gene (com um cromossomo X normal e um com o gene da hemofilia). Considerando que seu pai não é afetado, ele só pode passar um cromossomo X normal. Portanto, a mulher tem 50% de chance de ser portadora do gene da hemofilia (herdado da mãe). Assim, definimos a distribuição a priori do estado da mulher (portadora ou não) como sendo igual para ambas as possibilidades:

- $\Pr(\theta = 1) = 0.5$ (probabilidade de ser portadora)
- $\Pr(\theta = 0) = 0.5$ (probabilidade de não ser portadora)

Modelo de Dados e Verossimilhança. Considerando que a mulher tem dois filhos e nenhum dos dois é afetado pela hemofilia, podemos usar essa informação para atualizar nossa crença sobre o estado genético dela. Supondo que os filhos não sejam gêmeos idênticos, os eventos (estados de saúde dos filhos) são independentes, dado o estado genético da mãe. As probabilidades condicionais, ou verossimilhanças, são calculadas como segue:

- Se a mulher é portadora ($\theta = 1$), cada filho tem 50% de chance de não herdar o gene defeituoso e ser não afetado. Assim, a probabilidade de ambos os filhos serem não afetados é:

$$\Pr(y_1 = 0, y_2 = 0 | \theta = 1) = 0.5 \times 0.5 = 0.25.$$

- Se a mulher não é portadora ($\theta = 0$), ela não possui o gene para passar adiante, então a probabilidade de ambos os filhos serem não afetados é certa:

$$\Pr(y_1 = 0, y_2 = 0 | \theta = 0) = 1 \times 1 = 1.$$

Na verdade, existe uma probabilidade não nula de ser afetado mesmo que a mãe não seja portadora, mas esse risco – a taxa de mutação – é pequeno e pode ser ignorado para este exemplo.

Distribuição a posteriori. Após observarmos que ambos os filhos da mulher não são afetados pela hemofilia, podemos utilizar a regra de Bayes para atualizar nossa crença sobre se ela é portadora do gene. Usando y para denotar os dados conjuntos (y_1, y_2) , podemos usar a [Equação 1.1](#) e então,

$$\Pr(\theta = 1 | y) = \frac{p(y | \theta = 1) \Pr(\theta = 1)}{p(y | \theta = 1) \Pr(\theta = 1) + p(y | \theta = 0) \Pr(\theta = 0)} = \frac{(0.25)(0.5)}{(0.25)(0.5) + (1.0)(0.5)} = 0.20.$$

Isso indica que, com a informação de que ambos os filhos não são afetados, a probabilidade da mulher ser portadora do gene da hemofilia diminui para 20%. Isso é intuitivo, pois a ausência da doença nos filhos sugere que ela possa não ter o gene para transmitir.

Adicionando Mais Dados. A análise Bayesiana facilita a atualização sequencial das crenças com a adição de novos dados. Se a mulher tiver um terceiro filho que também não é afetado, não precisamos recomençar todo o cálculo. Em vez disso, a distribuição a posteriori obtida anteriormente se torna a nova distribuição a priori:

$$\Pr(\theta = 1 | y_1, y_2, y_3) = \frac{(0.5)(0.20)}{(0.5)(0.20) + (1)(0.8)} = 0.111.$$

Essa probabilidade mais baixa reflete o fato de mais uma evidência indicar que a mulher provavelmente não é portadora do gene. Por outro lado, se o terceiro filho fosse afetado, a probabilidade a posteriori de ela ser portadora aumentaria para 1, assumindo a ausência de mutações, o que significaria que ela definitivamente tem o gene, dado que a doença só aparece se o gene estiver presente.

1.4.2 Correção Ortográfica

RESOLUÇÃO DOS PROBLEMAS

Capítulo 1

1.1.

REFERÊNCIAS BIBLIOGRÁFICAS

- Bayes, M. and Price, M. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London Series I*, 53:370–418, January 1763.
- Bernshtein, S. N. An attempt to axiomatize the foundations of the theory of probability (in russian). *Soobshch. Khar'k. matem. ob-va*, 15:209–274, 1917.
- Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3-4):172–198, March 1927. doi: 10.1007/BF01397280.
- Jeffreys, H. *Theory of Probability*. International series of monographs on physics. Clarendon Press, 1998.
- Kolmogorov, A. *Foundations of the Theory of Probability*. Chelsea Publishing Company, 1950.
- Metropolis, N. and Ulam, S. The monte carlo method. *Journal of the American Statistical Association*, 44(247): 335–341, 1949. doi: 10.1080/01621459.1949.10483310.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 06 1953.
- Ross, S. *Introduction to Probability Models*. Elsevier Science, 2019.
- Siegrist, K. Random: Probability, mathematical statistics, stochastic processes. <https://www.randomservices.org/random/>, 2019.
- Thomson, J. J. Xl. cathode rays. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 44(269):293–316, 1897. doi: 10.1080/14786449708621070.
- Wilks, S. *Mathematical Statistics*. Princeton University Press, 1947.