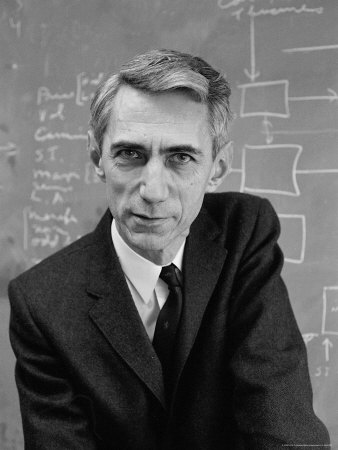
**Teoria da Informação**

**Entropia como medida de incerteza ou de informação:**



Claude Elwood Shannon (1916 — 2001) matemático e engenheiro eletrônico americano conhecido como "o pai da teoria da informação". Ver apêndice para uma biografia de Shannon.



Ludwig Bolzmann

**INCERTEZA:**

Para iniciar essa discussão precisamos dar um sentido matemático à palavra INFORMAÇÃO ou INCERTEZA e definir uma MEDIDA DA INFORMAÇÃO. Para isso o melhor é utilizar a metodologia axiomática na qual definimos certas propriedades que esperamos que a medida possua através dos axiomas e esperamos que os mesmos limitem a classe de funções capazes de expressar a medida, ou, no melhor dos casos, que exista apenas uma função que satisfaz aos axiomas. Claro que, nesse caso, todas as grandezas que satisfazerem aos mesmo axiomas serão agrupadas e chamadas pelo mesmo nome.

Então vamos começar esse capítulo explicitando um conjunto de axiomas que representem algo que definimos como informação. Suponha uma variável aleatória que pode tomar valores no conjunto finito  com probabilidades  com as restrições usuais  e . Queremos associar uma função de  que represente uma medida de incerteza associada à variável aleatória . Vamos construir duas funções  e . A função  será a incerteza associada a um evento de com probabilidade . Se o evento  tem probabilidade  então  é a incerteza associada ao evento . Podemos pensar que  é a incerteza removida ao descobrir que o evento  ocorreu, ou ainda, que é a informação revelada pelo fato de que  ocorreu. Já a função  é a inceretza média associada ao conjunto dos eventos , logo:



Dessa forma, é a incerteza média removida, ou informação média revelada, por descobrir o valor de . Com essas definições vamos impor axiomas sobre a grandeza INCERTEZA ou INFORMAÇÃO.

**Axioma 1:** Suponha que todos os valores de  são independentyes entre si e igualmente prováveis. Nesse caso  e  será uma função de . Por exemplo, no jogo de cara ou coroa de uma moeda honesta teremos , e no jogo de escolher uma pessoa ao acaso de uma cidade com  habitantes teremos . Aqui cabe a pergunta: o que contém mais informação – descobrir que se obteve coroa em um jogo de cara-ou-coroa ou descobrir que o cidadão José Jesus Pedro (também, com tanta ajuda dos santos) foi o escolhido em uma cidade de 20 milhões de habitantes? Claro que o evento de menor probabilidade, que causa maior surpresa, deve conter mais informação quando revelado. Esse é o primeiro axioma da teoria da informação:

 é monotônica crescente em .

Em outras palavras, se , então .

**Axioma 2:** Considere agora um experimento envolvendo duas variáveis aleatórias independentes. Uma denominada por  com valores possíveis  com iguais probabilidades , e a outra denominada  com valores possíveis  com iguais probabilidades . Como as variáveis são independentes então:



Agora, o fato de ter-se revelado o valor de  não traz qualquer informação sobre o valor de , pois as variáveis são independentes. Assim, a incerteza removida por descobrir o valor de  é apenas a inceretza média de . O mesmo argumento vale para . Assim, se sorteamos primeiro o valor de x extraímos a incerteza  e depois sorteamos  extraindo a incerteza , e no processo inteiro extraímos a incerteza . Agora se sorteamos as duas variáveis simultaneamente, temos 1 possibilidade em , logo extraímos a incerteza . O segundo axioma afirma que é indiferente a ordem de extração da informação, uma antes da outra ou as duas simultaneamente. Matematicamente isso é escrito como:



**Axioma 3:** Agora vamos remover a restrição de probabilidades iguais e pensar em termos de reagrupamento. Vamos agrupar a variável  em dois conjuntos  e  da seguinte forma:  e . Vamos sortear a variável  com o seguinte processo: primeiro sortear qual grupo,  ou , e sortear  dentro de seu respectivo grupo.

Agora, pelos axiomas da probabilidade:





Daí precisamos responder à pergunta: qual a probabilidade de sortear  sabendo que  foi escolhido? Para responder a essa pergunta usamos o teorema de Bayes:  (lê-se essa expressão como p de A dado B) onde . Se  então  e . Como  então  então:



Da mesma forma:



Vamos denominar o experimento de  o processo de escolher primeiro o grupo  e depois o , para diferenciá-lo do experimento escolher o  diretamente. A probabilidade será . Mas:

 e  logo 

como deveria ser, claro. Quanta incerteza será removida por descobrir qual dos dois grupos foi escolhido?



Dado que o grupo  foi escolhido qual a incerteza remanescente?



A mesma coisa para o grupo :



Assim a incerteza média removida no processo de escolha entre um grupo e outra será:



A incerteza na escolha de um dado  será a incerteza na escolha entre os dois grupos adicionada da incerteza média restante na escolha de  dentro de cada grupo, ou seja:



**Axioma 4:** Simplesmente impõe que  seja contínua em .

Os 4 axiomas da medida da incerteza então são:

1.  é monotônica crescente em.
2. .
3. 
4.  é contínua em .

Com esses 4 axiomas podemos provar o seguinte teorema:

A única função que satisfaz aos 4 axiomas é: . A base do logarítmo não interessa porque pode ser incorporada na constante arbitrária .

Provar que a função dada satisfaz aos axiomas é fácil.

1.  e é uma função monotônica crescente desde que a base seja um número mairo do que 1.
2.  e  logo  logo .
3. 







 CQD

1.  que é uma função contínua no intervalo .

O teorema entretanto afirma mais. Afirma que essa é a única função que satisfaz aos 4 axiomas. Podemos provar o teorema completo da seguinte forma:

1. Mostrar que  por indução.

Primeiro  logo  é verdadeiro. Supor que  é verdadeiro e provar que é verdade para . Pelo axioma 2 temos que  CQD.

1. Mostrar que  com .

Para , pelo axioma 2, temos que:  o que significa que . Logo  é verdadeira para . Para  vamos tomar um inteiro positivo  qualquer. Sempre haverá um . Para provar isso basta aplicar o  em todos os fatores, lembrando que para  então  é monotônica crescente, que nos leva a:

 ou ainda ,

então  é o antecessor de  e  o sucessor. Pelo axioma 1 sabemos que:



e pelo teorema (a) temos que , ou seja:

.

Por outro lado , logo, ou seja:



Podemos subtrair uma equação da outra considerando os piores casos tanto nos valores positivos quanto nos negativos. Do lado negativo o pior caso é , e no positivo é , assim:

 ou seja 

Como esse resultado é válido para qualquer , então podemos fazer  obtendo:

 ou que 

então:

 CQD

1. Provar que  se  é racional.

Se  é racional então  ode  e  são inteiros positivos e  porque . Agora podemos criar a função com  a partição:



Aplicando o axioma 3 temos:



aplicando (b) obtemos:



Da qual extraímos:



Portanto:

 CQD

Como essa função é válida para todo número racional e a função é contínua então o resultado é válido para qualquer .

1. Finalmente provar que 

Já sabemos que a fórmula é verdadeira para , , , e para , ,  e . Vamos supor que a fórmula é verdadeira para  e mostrar que, então, é verdadeira para . Para mostrar o último ponto precisamos do axioma 3.















Chegando ao resultado final



Ou seja:



Quando Shannon descobriu essa medida ele notou que era idêntica à entropia de Boltzmann:



e a denominou de ENTROPIA, hoje também conhecida como entropia de Shannon.

Se fizermos a constante  e usar o logaritmo na base 2 a incerteza é medida em unidades de bits = binary digits. Foi Shannon quem primeiro usou a palavra bit no paper de 1948, que ele atribuiu a John Tukey, seu colega no Bell Labs.

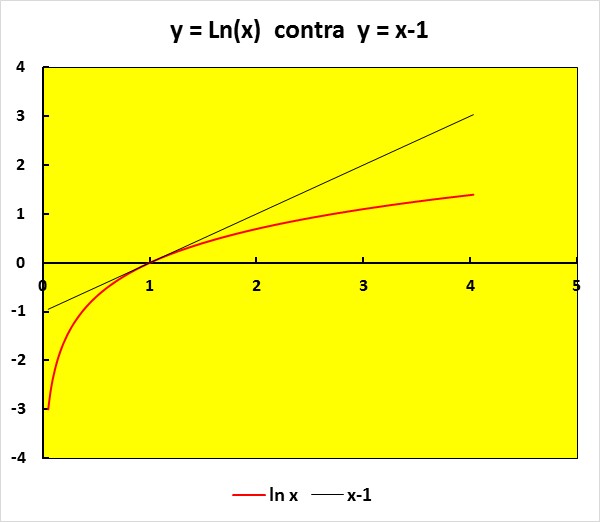
**Propriedades da Função Incerteza**

Lema 1: sejam  e  dois conjuntos de números positivos que obedecem à restrição . Então:



O logaritmo é uma função convexa, o que significa que a função está sempre abaixo da reta tangente em qualquer ponto. Agora:

 de modo que . Além disso  então a reta tangente no ponto  é dada por . Dessa forma sabemos que  e a igualdade só é válida em .



Isso significa que , portanto:



Logo  mas , ou seja, , multiplicando por -1 e invertendo a desigualdade obtemos , ou seja:



Teorema 1: .

Prova: basta fazer  no lema 1.



Portanto:



A igualdade só vale se .

**Incerteza Conjunta.**

Sejam  e  duas variáveis aleatórias associadas ao mesmo experimento. Seja . A função incerteza conjunta será:



Essa forma pode ser generalizada para  v.a.s.



**Teorema 2:** .

Prova:

 e 

então:





Da mesma forma:



Somando as duas:





Agora vale notar que



então a grandeza  representa um conjunto de probabilidades válido, tal que . Assim:



Portanto:



A igualdade só é válida se , ou seja, se as variáveis forem independentes.

**Incerteza Condicional.**

Definimos

 e 

como a média das incertezas condicionadas:







Agora  então:



Teorema: 

Prova:









Trocando  por  obtemos a outra parte

.

Teorema: 

Prova:

 logo

.

A igualdade é válida se as v.a.s  e  forem independentes.

**Medida da INFORMAÇÃO:**

A medida da informação é a medida da redução da incerteza. Assim, sem saber nada sobre a v.a. a incerteza era . Se soubermos  a incerteza diminui para . A informação ganha foi:







**Formalismo da Entropia Cruzada:**

Em lugar de maximizar a entropia podemos minimizar a informação. Vamos examinar o problema de achar as probabilidades de cada face de uma dado sabendo o valor médio das faces do dado, ou seja, o problema de inferir qual a distribuição de probabilidades  a partir de certas observáveis pode ser generalizado colocando o problema como:

Encontrar  que maximiza a entropia  com as restrições:



Que resolvemos através do Lagrangeano  obrigando que . Suponha que você tem um pouco mais de informação sobre o seu dado. Você sabe que ele veio da mesma fábrica de outro dado que foi totalmente caracterizado e que apresentou as probabilidades . Claro que se espera, então, que  e desejamos considerar essa informação extra no nosso processo de otimização. Assim é melhor resolver o seguinte problema:

Minimizar a informação:



Sujeito às restrições:

 e .

Note que se  é constante, então  e a informação se torna . Nesse caso nada muda pois somar uma constante não muda a maximização/minimização de uma função. Se  entretanto o problema é dado por:









Com as restrições:

 logo 

Junto com , logo  então ou ainda:



Assim estamos considerando a informação prévia de que .

Suponha agora que  de modo que , então  onde  , então:



logo:



Essa grandeza era já bem conhecida na aplicação do teste do  de Pearson [1857-1936] desenvolvido por volta de 1900. Esse teste mede o quão próximo da distribuição de  está da distribuição de  . Não é uma distância porque não obedece ao axioma de  uma vez que  mas tem a propriedade de .

**Entropia Cruzada de Cressie & Read:**



Noel A. Cressie

Vale a pena notar que é o termo  em  quem dá a “distância” entre as duas distribuições. Por outro lado se  então  também. Com isso Cressie e Read sugeriram uma outra medida de “distância” em 1984 dada por:



Para evitar problemas com vamos analisar o que ocorre no limite:

 que é do tipo  logo podemos aplicar L´Hopital, derivando em cima e embaixo em relação à :



Isso significa, então, que no limite de  caímos de volta na entropia cruzada anterior. Eles definiram então a seguinte entropia cruzada:

 para garantir que a função exista em  e que para ,  seja a mesma entropia cruzada anterior.

**Entropia de Renyi:**



Renyi sugeriu outra entropia em 1970:



Se  então  mas  e  logo temos, novamente, um limite do tipo  . Aplicando L´Hopital novamente:





Então:

 assim 

**Entropia de Tsallis:**



Constantino Tsallis nasceu em Atenas na Grécia, emigrou aos quatro anos para Mendoza na Argentina. Graduação e mestrado em Física pelo Instituto Balseiro de Bariloche da Universidad Nacional de Cuyo e concluiu o doutorado em Física pela Universidade de Paris Orsay. Chegou ao Brasil em 1975 imigrou para o Brasil, inicialmente trabalhando na Universidade de Brasília até se tornar professor titular do Centro Brasileiro de Pesquisas Físicas, onde permanece até hoje.

Tsallis sugeriu outra entropia em 1988 conhecida como entropia e Tsallis:



Novamente temos problemas quando  com limite do tipo porque  e . Aplicando L´Hopital:



Então .

Como Cressie & Read só definiram uma entropia cruzada vamos escrever as entropias cruzadas de Renyi e Tsallis:







As entropias de Renyi e Tsallis se tornam a entropia de Shannon/Boltzmann para  enquanto a de Cressie & Read para . Essas entropias estão relacionadas entre si. Vamos fazer  nas entropias de Renyi e Tsallis:



 então  portanto:



Por outro lado:

 logo:



Assim:





Agora todas essas entropias cruzadas tendem a entropia cruzada de Shannon/Boltzmann quando . Uma diferença importante entre elas se encontra na propriedade da aditividade. A propriedade  só é válida para Shannon/Boltzmann, e não para as outras três. Além disso se  e  são independentes então  vale para Shannon/Boltzmann e Renyi, mas não vale para Tsallis. Já o axioma (3) para a incerteza vale para Shannon e Tsallis mas não vale para Renyi.

Golan e Perloff examinaram as propriedades dessas entropias como estimadores para uso geral em estatística frente a 5 axiomas desejados para um bom estimador. Apenas o estimador baseado na entropia de Shannon/Boltzmann satisfaz aos 5 axiomas. Os outros dois deixam de satisfazer a um deles, Só recomenda o uso de outras entropias para problemas em que as matrizes obtidas pela entropia convencional apresentam dificuldades de cálculo, ou em amostras muito pequenas.

**Teoria da Informação, Entropia e Medida de Desigualdade:**

Vamos recolocar o problema da informação em outro contexto. Suponha uma população com  habitantes em que a iésima pessoa recebe a fração  da renda total do país  . Aqui  significa YIELD e representa o Produto Interno Bruto, também conhecido como PIB [GDP em inglês para Gross Domestic Product]. Isso significa que a renda do iésimo habitante vale . Assim percebemos que , pois , e que , logo  é uma probabilidade legítima. A incerteza da distribuição de renda é dada por:



Será máxima quando  , ou seja, uma distribuição de renda totalmente igualitária. Dessa forma . A função  só pode ser nula se toda a renda do país ficar com uma pessoa apenas e nada para todo o restante da população [significaria a destruição daquela sociedade], e será máxima, , se for igualmente distribuída, com cada habitante recebendo a fração  da renda total. Isso sugere que se poderia usar a função  como uma medida da igualdade/desigualdade da distribuição de renda. Entretanto, ela é uma função que é máxima na igualdade total e que depende do número de habitantes, o que não é bom para uma medida de desigualdade, que deve crescer com a desigualdade. Assim, o número:



É bom indicador da desigualdade. Na economia e teoria da informação a unidade usada para  é o bit e para  é chamada NIT, N de natural ou neperiano. Esse índice é conhecido como índice T de Theil apresentado pelo holandês Henri Theil [1924-2000] em 1967 na publicação “Economics and Information Theory”.



Henri Theil

Agora notamos que esse índice é uma entropia cruzada pois:



Ou seja



com , a distribuição real de renda, e , a distribuição igualitária de renda. Além de definir o índice  de Theil ele também definiu o índice  de Theil, dado por:





Vamos re-expressar esses índices em termos da renda , onde  é a renda média. Assim:





Já o índice  é dado por:







**Propriedades de agrupamento dos índices de Theil:**

Suponha que queremos agrupar os  habitantes em certas classes, como agrupamento por regiões de um país, por exemplo. Seja o késimo grupo com  habitantes e renda  e  . Vamos chamar a fração da população no grupo  através da letra . Note que:



Agora vamos somar e subtrair o termo:







Agora  é a fração da renda do iésimo habitante dentro do grupo  logo o índice  de Theil dentro desse grupo será . Por outro lado o índice  de Theil entre as regiões é dado por . Assim:



Para o índice  de Theil:



Agora:

 uma vez que  .









Novamente, reconhecemos:

 e  de modo que:



Assim, comparando os dois índices, vemos que:





Assim percebemos que o índice  é uma média entre grupos ponderada pela fração da renda enquanto o índice  é ponderado pela fração da população. Por isso se diz que o índice  é democrático. Se os grupos fossem quebrados por faixas de renda o índice  é mais sensível a alterações na desigualdade de renda mais alta e o  do grupo de renda mais baixa, na qual  é maior.

**Transferência de Renda:**

Dizemos que uma transferência de renda é progressiva quando ele se dá dos mais ricos para os mais pobres, e regressiva no caso oposto, quando os mais pobres transferem renda para os mais ricos. Ambos os índices aumentam com uma transferência regressiva de renda, pois a desigualdade aumenta. Podemos analisar o comportamento dos dois índices analisando apenas a transferência de renda entre duas pessoas, supondo que a distribuição de renda em todo o resto da população se manteve constante.

Vamos supor que dois habitantes, o 1, mais pobre, e o 2, mais rico. No grupo dos dois o habitante 1 tinha a participação  e o 2 a participação . Agora o habitante 1 perdeu parte de sua renda  para o habitante 2, em uma transferência regressiva de renda. Nesse caso as novas participações na renda do grupo se tornam:

 logo: .

Da mesma forma , então . Vamos chamar  e analisar a mudança da transferência de renda no índice :



A derivada do índice em função da transferência é dada por:





Agora, por hipótese  , então  logo  e . Assim , o que significa que a função  é crescente em . Tirar dos pobres para dar aos ricos aumenta o índice, e o inverso diminui o índice.

No índice  teremos:











Claro que  também

**Índice de Desigualdade de Atkinson:**



Em 1970 Sir Anthony Barnes Atkinson criou um conjunto de medidas de desigualdade através de uma função bem estar social. A ideia foi que seria possível criar uma função bem estar social somando as funções bem estar de cada um dos indivíduos. Supõe que as pessoas só se preocupam com suas próprias rendas, de modo que a renda do vizinho nem aumenta o bem estar próprio, caso de solidariedade, nem o diminui, caso da inveja. Assim considerou a função bem estar social (wellfare) dada por:



Onde  é a renda do iésimo indivíduo e a função  é a mesma para todas as pessoas. Admite-se que  é uma função crescente e côncava, ou seja, o acréscimo de bem estar de uma variação de renda  é menor para quem tem renda mais alta. Se uma função é côncava então ela está acima da reta secante, ou seja:



Com a igualdade sendo válida apenas no caso  . Para  temos que:



Esse resultado pode ser generalizado para:



Onde a igualdade só vale se . A renda média dessa sociedade é dada por:



A pergunta de Atkinson foi: Qual seria o nível de renda igualitária (mesma renda para todos)  que traria a mesma função bem estar social ? Ou seja:



Assim:



Como  é crescente e côncava, então:



Então  com a igualdade valendo apenas na distribuição igualitária  .

Assim Atkinson criou um índice que varia de zero a um, com zero para a distribuição igualitária e 1 para toda a renda concentrada em um habitante apenas:



Uma função crescente e côncava tem derivada positiva e decrescente. Assim vamos supor que:

 .

Nesse caso:



Assim:









O índice de Atkinson, portanto, será dado por:



Ou seja:



Tirando o logaritmo de ambos os lados:







Que é uma entropia cruzada de Renyi.

**Apêndice. Biografia de Claude E. Shannon.**



De 1932 a 1936, estudou matemática e engenharia elétrica na University of Michigan. Em 1948, publicou o importante artigo científico intitulado A Mathematical Theory of Communication enfocando o problema de qual é a melhor forma para codificar a informação que um emissor queira transmitir para um receptor. Neste artigo, trabalhando inclusive com as ferramentas teóricas utilizadas por Norbert Wiener, Claude Shannon propôs com sucesso uma medida de informação própria para medir incerteza sobre espaços desordenados.

Em 1949, em co-autoria com o também matemático estadunidense Warren Weaver (1894-1978), publicou o livro Teoria Matemática da Comunicação (The Mathematical Theory of Communication), contendo reimpressões do seu artigo científico de 1948 de forma acessível também a não-especialistas - isto popularizou seus conceitos.

Entre 1946 e 1953, Claude Shannon integrou temporariamente o grupo reunido sob o nome de Macy Conferences, contribuindo para a consolidação da teoria cibernética junto com outros cientistas renomados como John von Neumann e Norbert Wiener. Shannon é famoso por ter fundado a teoria da informação com um artigo publicado em 1948. Mas a ele também é creditado como fundador tanto do computador digital como do projeto de circuito digital em 1937, quando, com 21 anos de idade e aluno de mestrado no MIT, ele escreveu uma tese demonstrando que uma aplicação elétrica utilizando álgebra booleana poderia resolver qualquer problema de lógica. Tem sido dito que esta foi a tese de mestrado de mais importância de todos os tempos. Shannon contribui para o campo da criptoanálise durante a segunda guerra mundial.

Biografia

Shannon nasceu em Petoskey, Michigan. Seu pai, Claude Sr (1862–1934), um descendente dos primeiros colonos de New Jersey, foi um empresário bem sucedido e foi juiz por um certo tempo. Sua mãe , Mabel Wolf Shannon (1890–1945), filha de imigrantes alemães, era uma professora de línguas. Os primeiros 16 anos de Shannon foram em Gaylord, Michigan, onde ele frequentou o ensino público, graduando-se no Gaylord High School em 1932. Shannon mostrou uma inclinação para coisas mecânicas, seus melhores talentos eram para a ciência e matemática. Em casa construiu dispositivos tais como modelos de aviões, um modelo de um barco controlado por rádio e um sistema de telégrafo. Enquanto crescia, trabalhava como mensageiro da Western Union. Seu herói de infância era Thomas Edison, que descobriu depois ser um primo distante. Ambos eram descendentes de John Ogden, um líder colonial e um ancestral de muitas pessoas ilustres.

Teoria Booleana

Em 1932 Shannon começou a cursar a University of Michigan, formando em 1936 com duas graduações de bacharelado em engenharia elétrica e matemática. Posteriormente, começou seus estudos de pós-graduação no Massachusetts Institute of Technology (MIT), onde trabalhou com o analisador diferencial de Vannevar Bush, um computador lógico.

Ao estudar os complexos circuitos ad hoc do analisador diferencial, Shannon viu que os conceitos de George Boole, inventor da álgebra booleana, poderia ser útil para várias coisas. Um documento eleborado a partir da sua tese de mestrado em 1937, A Symbolic Analysis of Relay and Switching Circuits, foi publicado na edição de 1938 da Transactions of the American Institute of Electrical Engineers. Howard Gardner, da universidade de Harvard, chamou a tese de Shannon como "possivelmente a mais importante e também a mais famosa tese de mestrado do século."

Neste trabalho, Shannon provou que a álgebra booleana e a aritmética binária poderiam ser utilizadas para simplificar o arranjo dos relés eletromecânicos e então utilizados em comutadores para roteamento telefônico. Expandindo o conceito ele também mostrou que deveria ser possível a utilização de arranjos de relés para resolver problemas de álgebra booleana. A exploração dessa propriedade de interruptores elétricos criou a lógica e os conceitos mais básico dos computadores digitais. O trabalho de Shannon tornou-se o principal na área de circuitos digitais quando se tornou amplamente conhecido entre a comunidade de engenharia elétrica durante e após a segunda guerra mundial. O trabalho teórico rigoroso de Shannon substituiu completamente os métodos ad hoc que haviam prevalecido anteriormente.

Em 1940, Shannon se tornou pesquisador do Instituto nacional de Estudos Avançados em Princeton, Nova Jersey. Em Princeton, Shannon teve a oportunidade de discutir suas idéias com cientistas e matemáticos influentes como Hermann Weyl e John von Neumann, além de um encontro ocasional com Albert Einstein. Shannon trabalhou livremente em todas as áreas, e começou a moldar as idéias que se tornariam a teoria da informação.

**Pesquisa em Tempo de Guerra**

Shannon em seguida juntou-se a Bell Labs para trabalhar em sistemas de controle de artilharia e criptografia durante a segunda guerra mundial, sob um contrato com a seção D-2 (Seção de Controle de Sistemas) do Comitê Nacional de Pesquisa em Defesa.

Conheceu sua esposa quando era um analista numerico na Bell Labs. Casaram em 1949.

Durante dois meses no início de 1943, Shannon entrou em contato como o criptoanalista líder e matemático britânico Alan Turing. Turing havia sido enviado para Washington para compartilhar com o serviço de criptoanálise da marinha dos EUA os métodos utilizados pela escola de códigos e cifras do governo britânico em Bletchley Park para quebrar as cifras utilizadas pelos submarinos alemães no Atlântico Norte. Ele também ficou interessado em cifragem de fala e para isso ficou um tempo no Bell Labs. Shannon and Turing se encontraram na hora do lanche em uma cafeteria e Turing mostrou a Shannon seu artigo que definiu o que hoje é conhecido como a "Máquina Universal de Turing". Em 1945, quando a guerra estava chegando ao fim, O NDRC estava emitindo um resumo dos relatórios técnicos como sua última atividade antes de seu eventual fechamento. Dentro do volume de controle de fogo um documento especial entitulado "Suavização de Dados e Previsão em Sistemas de Controle de Fogo",coautoria de Shannon, Ralph Beebe Blackman, e Hendrik Wade Bode, formalmente tratava do problema de suavização dos dados no controle de incêndio por analogia com "o problema de separar um sinal de um ruído interferindo no sistema de comunicação." Em outras palavras foi modelado o problema em termos de dados e processamento de sinal e assim, anunciava o início da era da informação.

Seu trabalho em criptografia foi mais estreitamente relacionada com suas publicações posteriores sobre a teoria da informação. No final da guerra, ele preparou um memorando para a Bell Telephone Labs entitulado "Uma Teoria Matemática da Criptografia," datada de setembro de 1945. Uma versão desclassificada deste trabalho foi posteriormente publicada em 1949 como "Teoria da Comunicação de Sistemas Secretos" no Bell System Technical Journal. Este trabalho incorporou muitos dos conceitos e formulações matemáticas que também apareceram em seu "Uma Teoria Matemática da Comunicação". Shannon disse que suas idéias em teoria da comunicação e criptografia durante a guerra haviam sido desenvolvidas simultaneamente e " elas estavam tão juntas que você não podia separa-las" . Em note de rodapé perto do início do relatório classificado, Shannon anunciou sua intenção de "desenvolver estes estudos... em um memorando sobre a transmissão de informações."

Enquanto na Bell Labs, ele provou que a one-time pad era inquebrável em sua pesquisa que mais tarde foi publicada em outubro de 1949. Ele também provou que qualquer sistema inquebrável deve ter essencialmente as mesmas características do one-time pad: A chave deve ser verdadeiramente aleatória, tão grande quanto o texto original, nunca reutilizada e mantida em segredo.

**Contribuições Pós-Guerra**

Em 1948, o memorando prometido apareceu como "A Mathematical Theory of Communication", um artigo com duas partes nas edições de julho e outubro do Bell System Technical Journal. Este trabalho enfoca no problema da melhor forma de codificar uma informação que um remetente deseja transmitir. Neste trabalho fundamental ele usou ferramentas da teoria da probabilidade, desenvolvidas por Norbert Wiener, que estavam em seus estágios iniciais de serem aplicadas a teoria das comunicações na época. Shannon desenvolveu a entropia da informação como uma medida de incerteza em uma mensagem.

Contribuição fundamental da teoria da informação para processamento de linguagem natural e lingüística computacional foi ainda estabelecida em 1951, em seu artigo "Previsão e Entropia de Impresso Inglês", mostrando limites superior e inferior da entropia nas estatísticas de Inglês - dando uma base estatística para análise da linguagem.

Outro papel notável publicado em 1949 é a "Communication Theory of Secrecy Systems", uma versão desclassificada do seu trabalho em tempo de guerra sobre a teoria matemática de criptografia , no qual ele provou que todas as cifras teoricamente inquebrável deve ter os mesmos requisitos que a one-time pad. A ele também é creditado a introdução da teoria da amostragem, que se preocupa com o que representa um sinal de tempo contínuo a partir de um conjunto (uniforme) discreto de amostras. Essa teoria foi essencial para permitir a passagem das telecomunicações dos sistemas analógicos para sistemas digitais no ano de 1960 e posteriores.

**Hobbies e Invenções**

Fora de suas pesquisas acadêmicas, Shannon estava interessado em malabarismo , monociclo e xadrez. Ele também inventou diversos dispositivos. Um dos seus dispositivos mais engraçados era uma caixa mantida em sua mesa chamada de "Máquina definitiva", baseada em uma idéia de Marvin Minsky. Além disso, ele construiu um dispositivo que poderia resolver o Cubo de Rubik. Ele também é considerado o co-inventor do primeiro computador portátil, juntamente com Edward O. Thorp. O dispositivo foi utilizado para melhorar as chances quando se jogar roleta .

**Legado e homenagens**

Shannon chegou ao MIT em 1956, para se juntar ao corpo docente e para realizar um trabalho no Laboratório de Pesquisa de Eletrônica (RLE). Ele continuou a servir no corpo docente do MIT até 1978. Para comemorar suas conquistas, houve celebrações de seu trabalho em 2001, e atualmente há seis estátuas de Shannon esculpido por Eugene L. Daub : um na Universidade de Michigan , uma no MIT no Laboratório de Sistemas de Informação e Decisão , um em Gaylord, Michigan, um na Universidade da Califórnia, San Diego , uma no Bell Labs , e outro no Labs Shannon AT & T. Após o rompimento na Bell, a parte do Bell Labs que ficou com a AT & T foi nomeado Shannon Labs em sua honra.

De acordo com Neil Sloane , a perspectiva introduzida por Shannon da teoria da comunicação (agora chamada de teoria da informação ) é a base da revolução digital, e cada dispositivo que contém um microprocessador ou microcontrolador é um descendente conceitual da publicação de Shannon ".. Ele é um dos grandes homens do século, sem ele, nenhuma das coisas que conhecemos hoje existiria. Toda a revolução digital começou com ele ".

Shannon desenvolveu a doença de Alzheimer , e passou seus últimos anos em um asilo de Massachusetts. Ele foi auxiliado por sua esposa, Mary Elizabeth Moore Shannon, o filho, Andrew Moore Shannon; a filha, Shannon Margarita, uma irmã, Catherine S. Kay e suas duas netas.

**Apêndice. Generalização do Método de Entropia Máxima.**

O problema de inferir qual a distribuição de probabilidades  a partir de ertas observáveis pode ser generalizado colocando o problema como:

Encontrar  que maximiza a entropia

Maximizar  com as restrições:



O Lagrangeano do sistema é dado por:





, . Definindo , , , ... e  as  constantes devem ser extraidas do sistema de equações não lineares das restrições:



Que pode ser re-escrito da forma:



Nesse caso, então, a estratégia é: usar SOLVER para encontrar com as equações:



O valor de é então calculado com .

Caso particular com apenas uma restrição de valores médios:

O Lagrangeano do sistema é dado por:





Logo  onde . As duas restrições são:

1. 

2. 

 logo  define o valor de . Com esse valor temos que .

Condições para existência de solução. Note que  então se todos os  ou todos os  não haverá solução para a equação. Seja  e , logo é preciso, então, que .



