Statistical Mechanics II

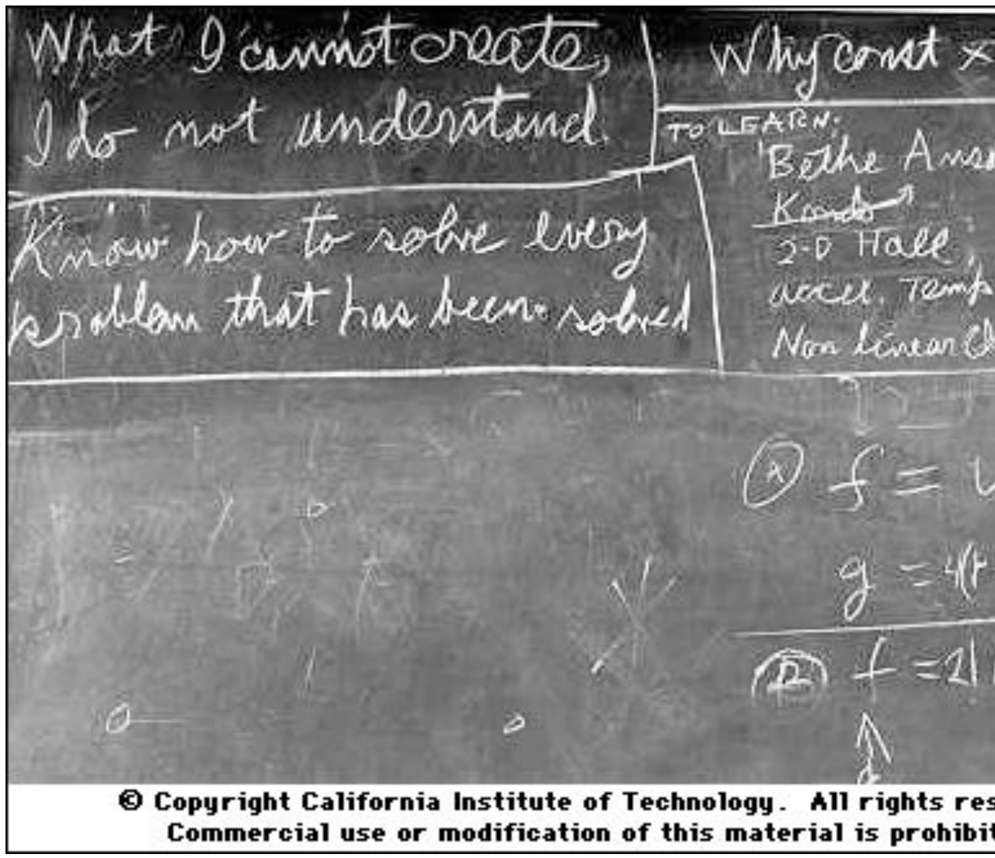
Jiansheng Wu

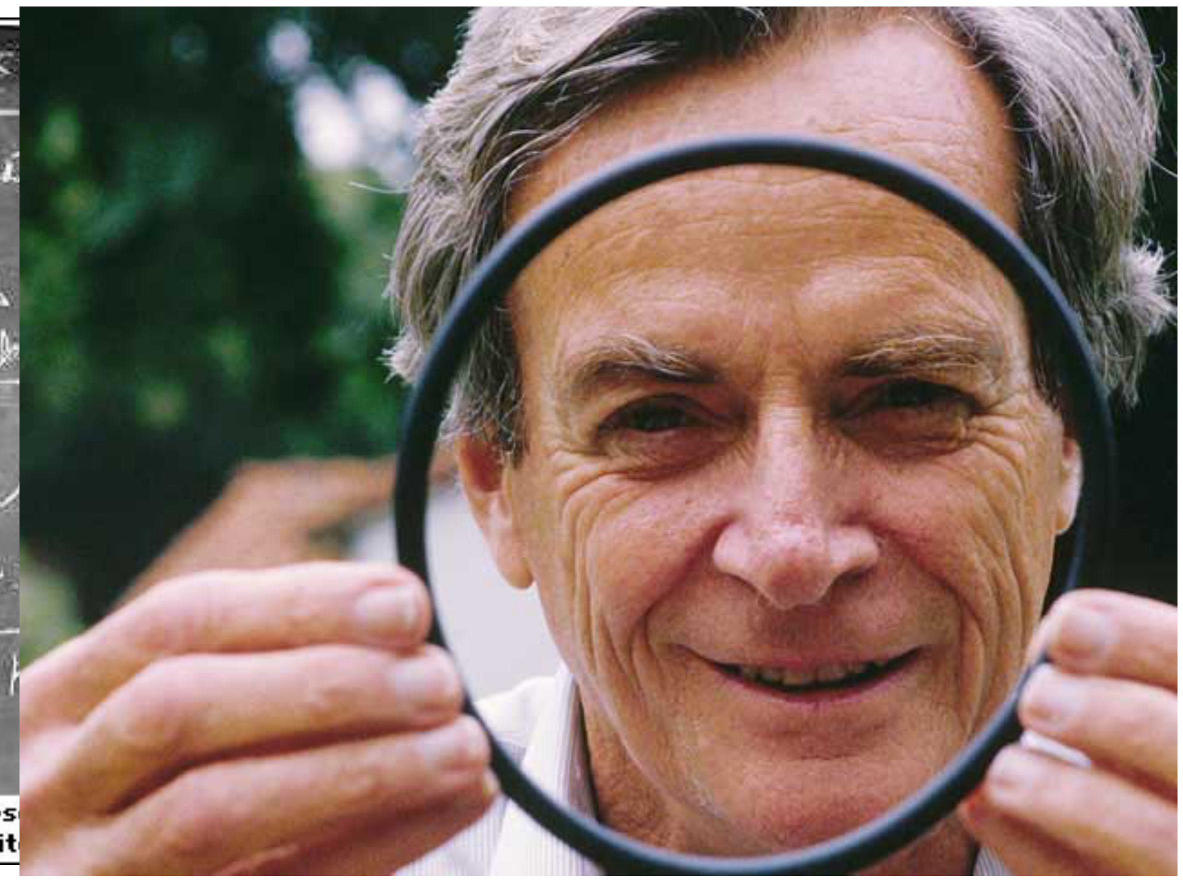
Lecture 2

Outline

- Probability
- Rule of large number
- Entropy

About Homework

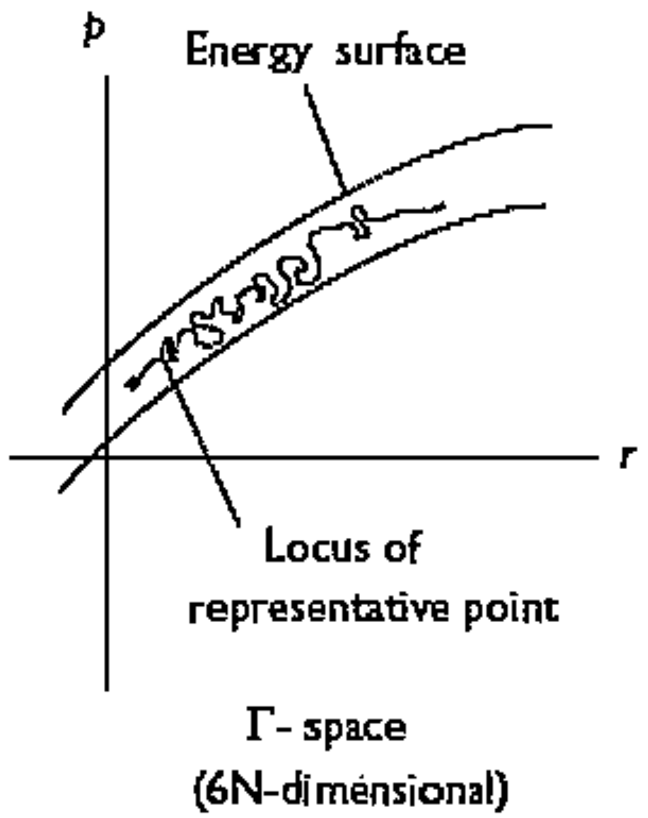




He meant that, starting with a blank piece of paper and the knowledge already in his mind, he could take any theoretical result and re-derive it.

Why do we need probability in SM?

- Two methods to get the thermodynamical properties
- 1) Kinetic theory: using the equation of motion of the system to calculate the properties
- 2) Ensemble theory: ignoring the actual dynamics, using the distribution. (So we need probability)

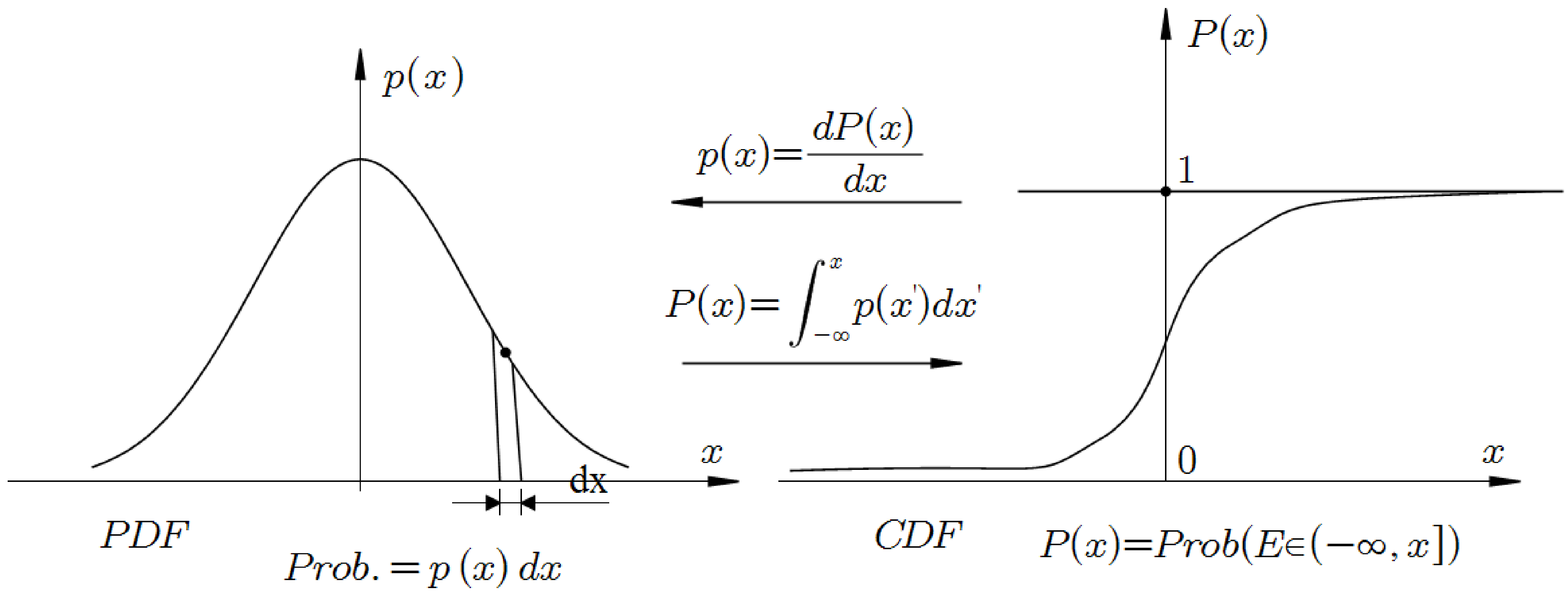


Statistics made simple!

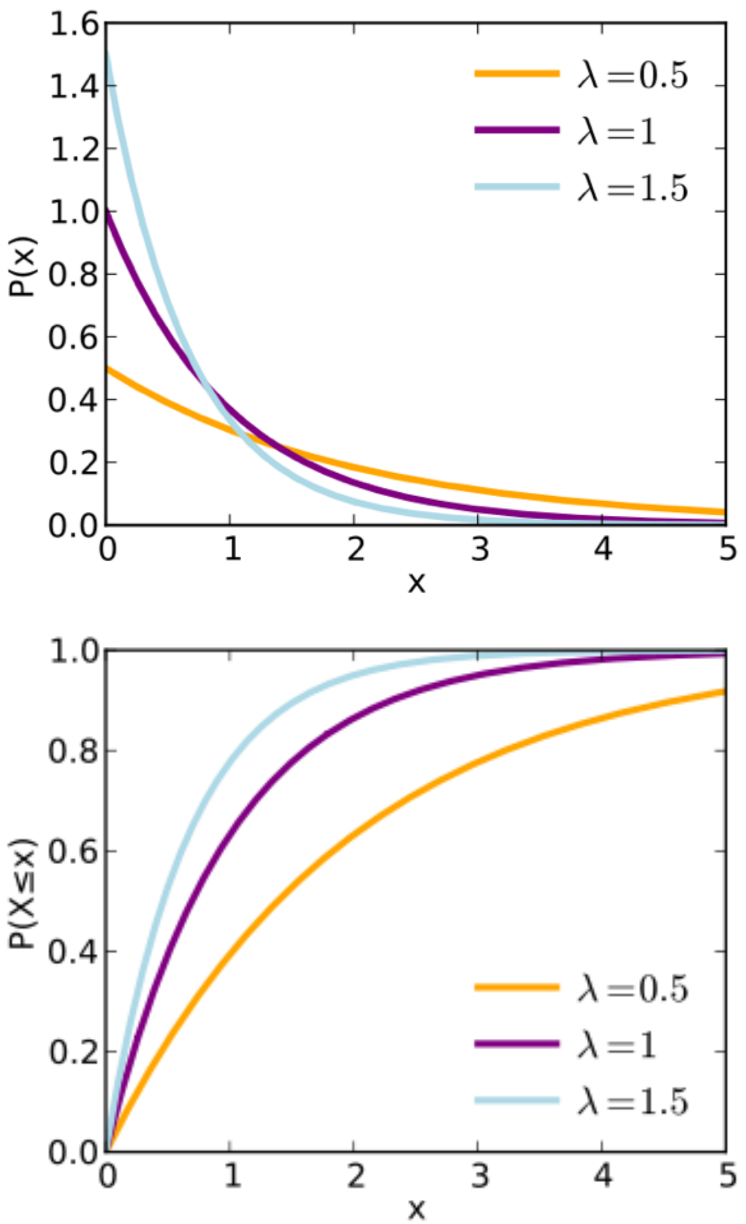
Probability

probability density function（PDF）概率密度函数

cumulative probability function（CPF）累积概率函数

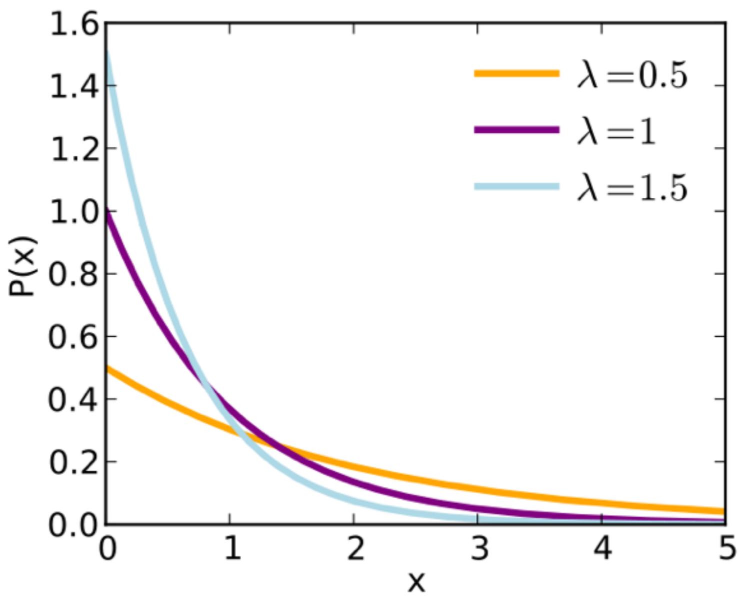


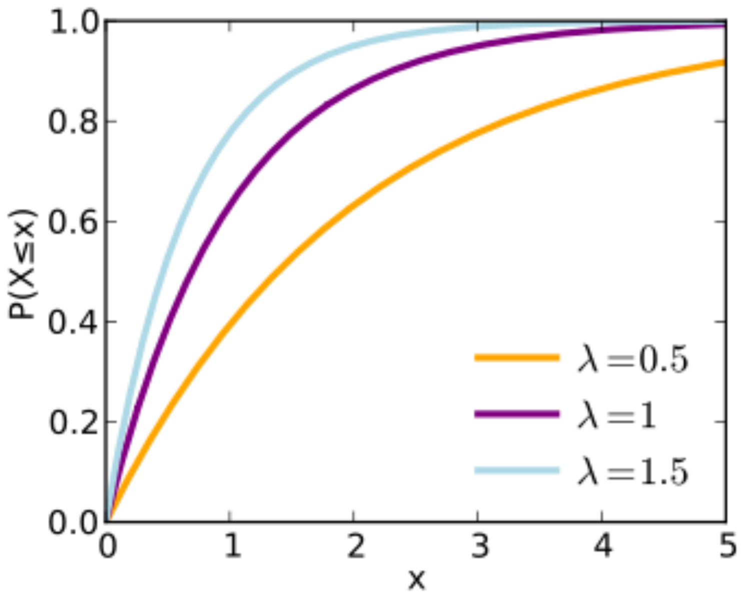
Exponential distribution



Parameters	$\lambda > 0$ rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$
Mean	λ^{-1} $(= \beta)$
Variance	λ^{-2} $(= \beta^2)$
Entropy	$1 - \ln(\lambda)$

Exponential distribution

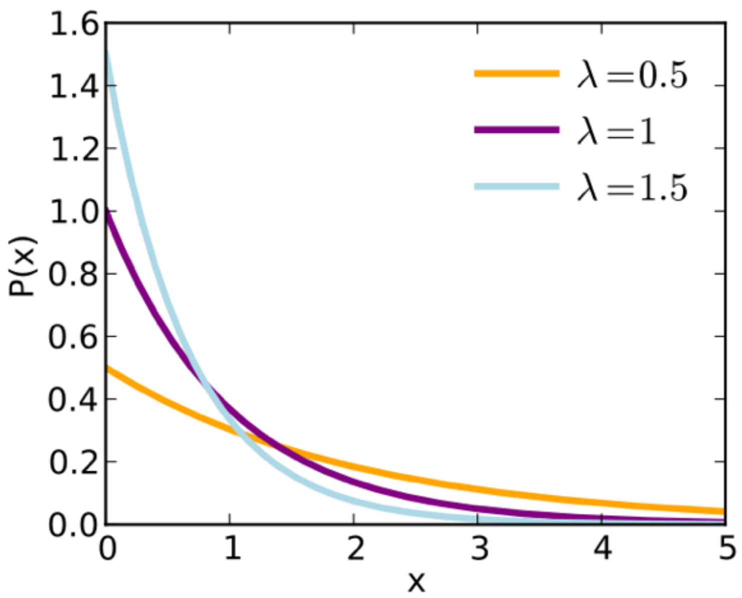


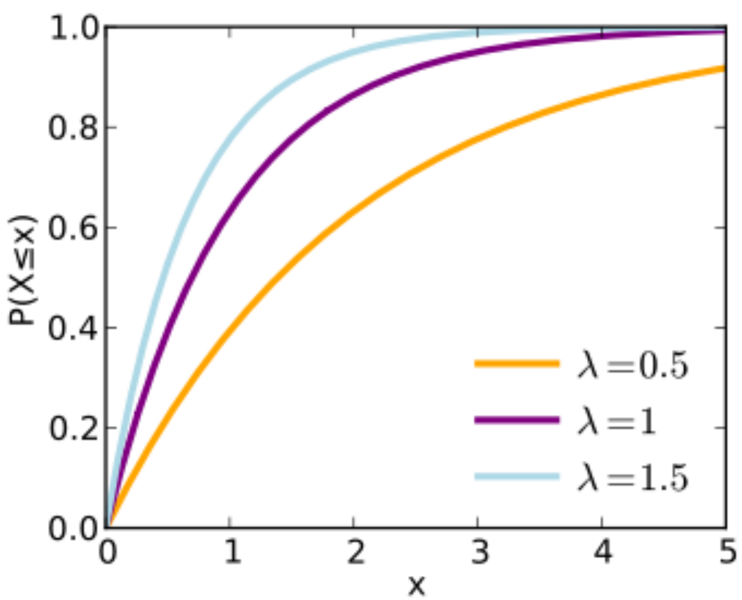


Parameters	$\lambda > 0$ rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$
Mean	λ^{-1} $(= \beta)$
Variance	λ^{-2} $(= \beta^2)$
Entropy	$1 - \ln(\lambda)$

$$\mathrm{E}(x) = \int_{-\infty}^{\infty} x p(x) \, dx$$

Exponential distribution





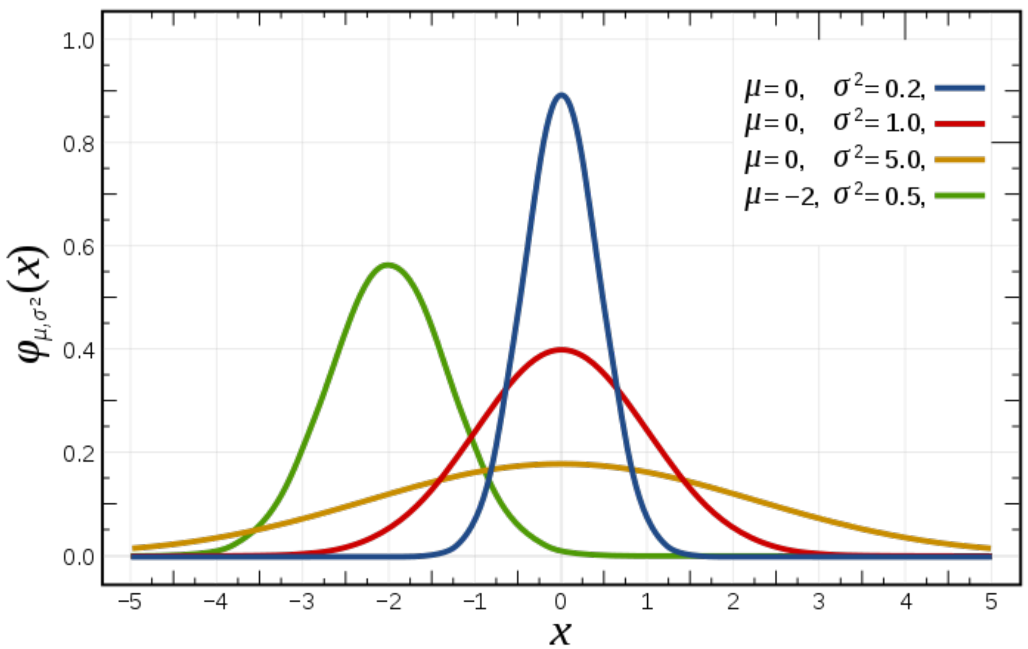
Parameters	$\lambda > 0$ rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$
Mean	λ^{-1} (= β)
Variance	λ^{-2} (= β^2)
Entropy	$1 - \ln(\lambda)$

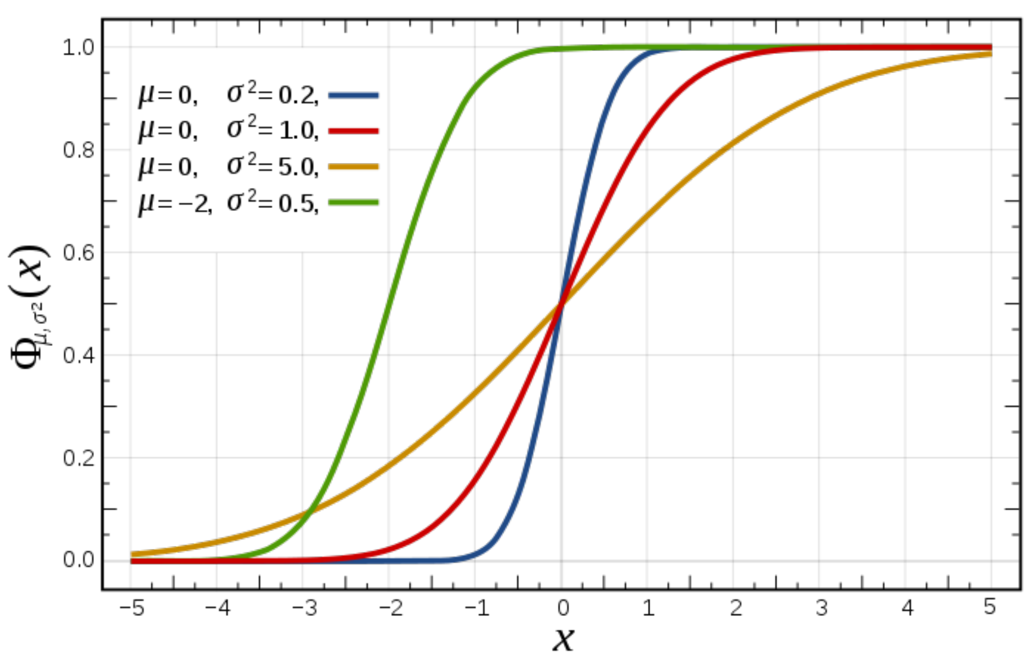
$$E(x) = \int_{-\infty}^{\infty} x p(x)\, dx$$

$$\begin{aligned} Var[x] &= E\left[(x - E[x])^2\right] \\ &= E\left[x^2\right] - (E[x])^2 \end{aligned}$$

$$\sigma = \sqrt{Var[x]}$$

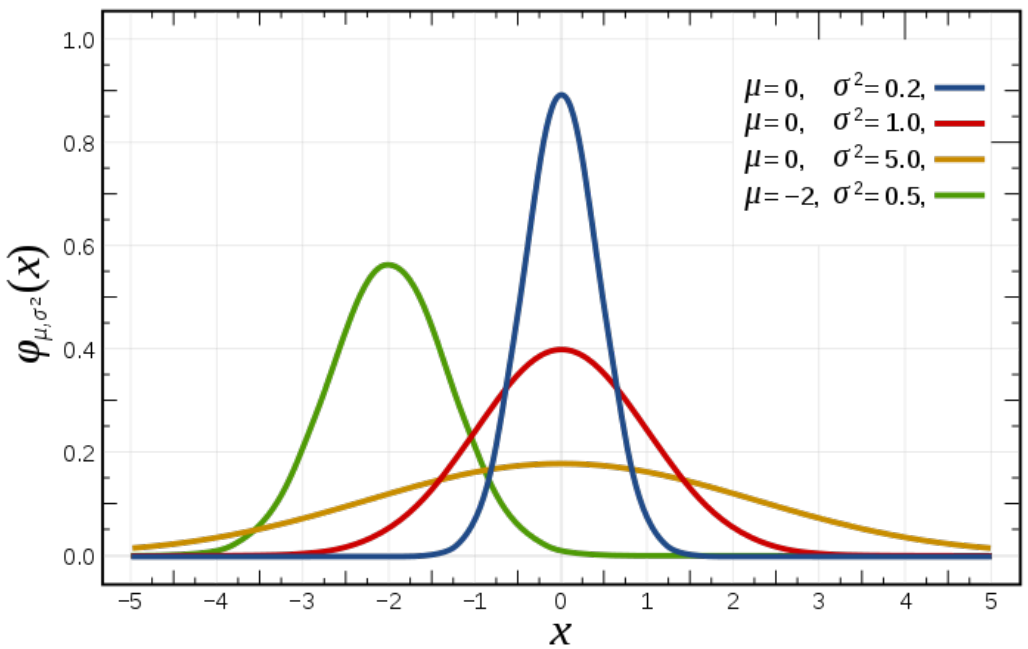
Gaussian distribution

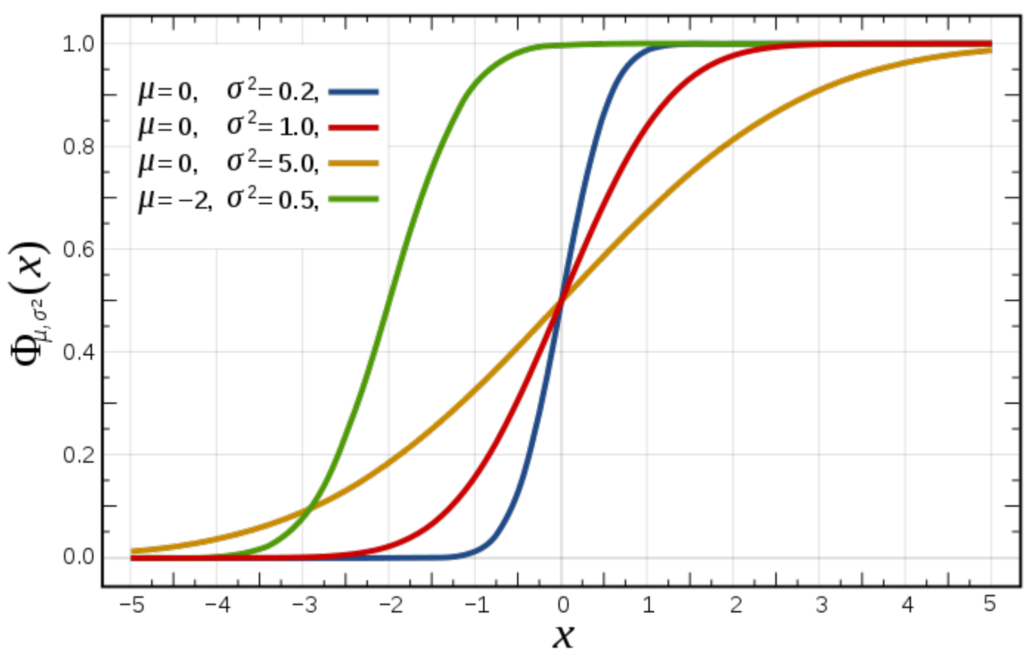




Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Mean	μ
Variance	σ^2
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$

Gaussian distribution



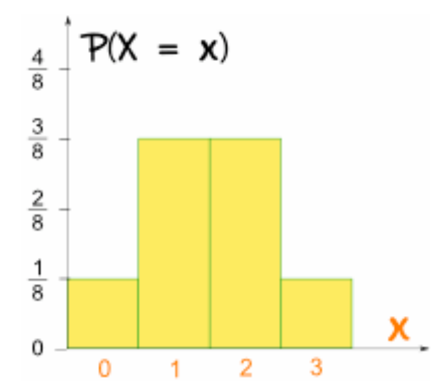


Notation	$\mathcal{N}(\mu,\sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Mean	μ
Variance	σ^2
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$

Binomial distribution

Tossing a coin, we get “head” (H) and “tails” (T)
the probability is all p=0.5

Tossing N=3 times, what is the probability to get n=2 H?



It is symmetrical!

number of configuration with 2 H in 3 tossing

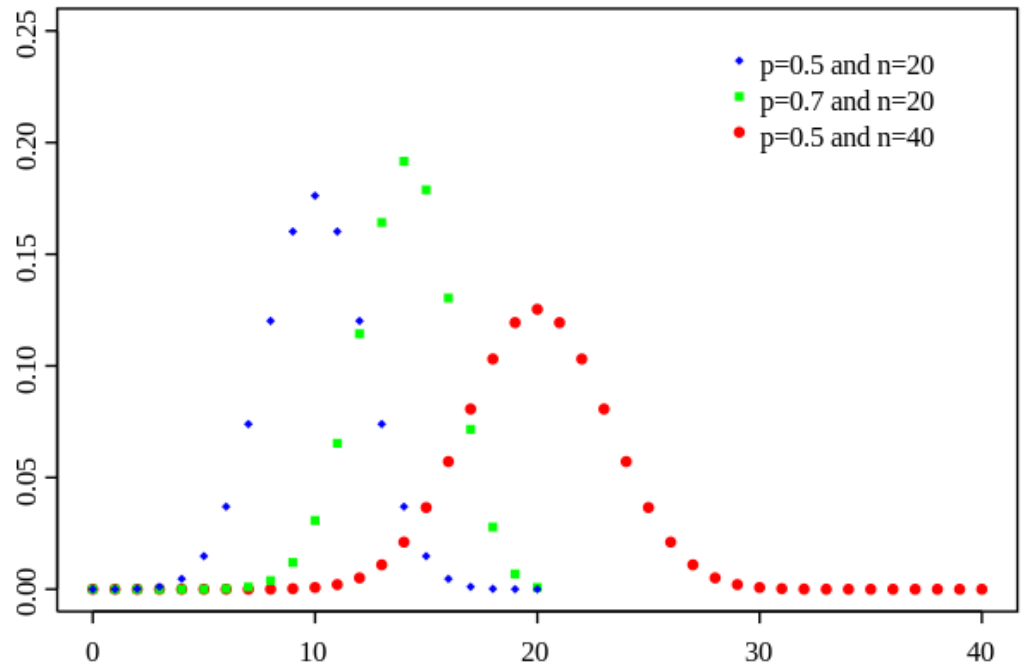
$$\binom{3}{2}(0.5)^2(0.5)^1 = 3/8$$

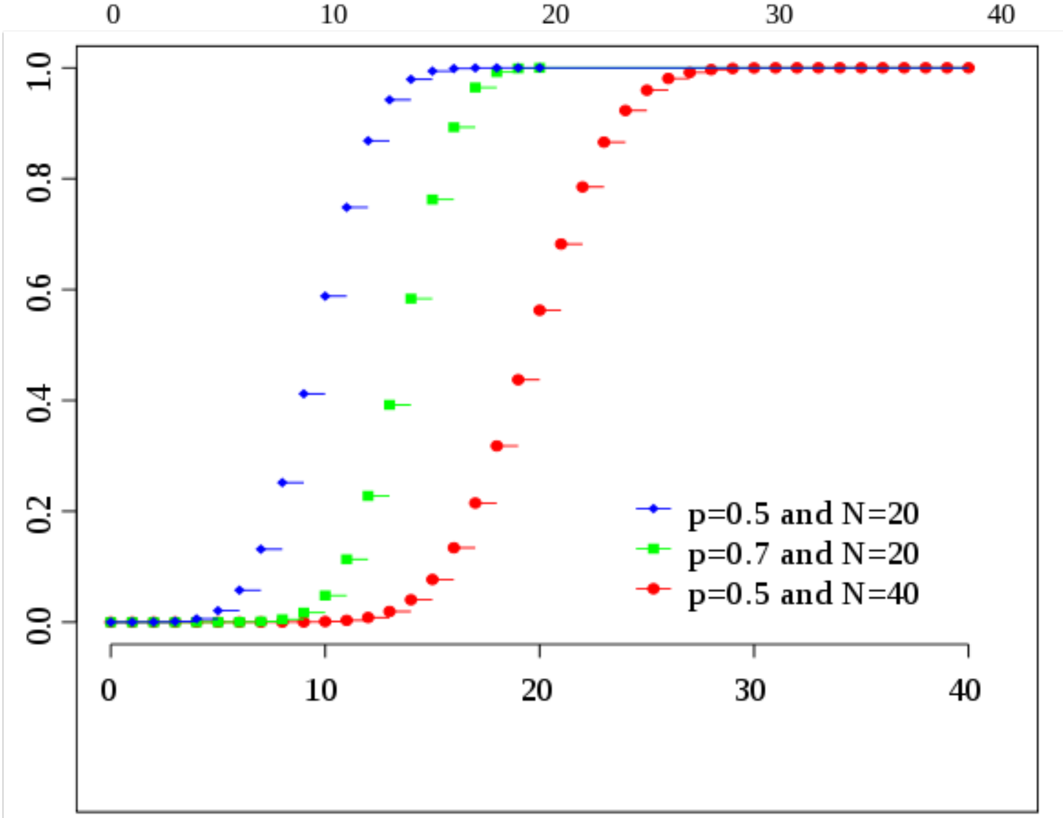
probability of 2 H

probability of 1 T

- P(Three Heads) = P(**HHH**) = **1/8**
- P(Two Heads) = P(**HHT**) + P(**HTH**) + P(**THH**) = 1/8 + 1/8 + 1/8 = **3/8**
- P(One Head) = P(**HTT**) + P(**THT**) + P(**TTH**) = 1/8 + 1/8 + 1/8 = **3/8**
- P(Zero Heads) = P(**TTT**) = **1/8**

Binomial distribution





Notation	$B(n, p)$
Parameters	$n \in \mathbf{N}_0$ — number of trials $p \in [0,1]$ — success probability in each trial
Support	$k \in \{0, \ldots, n\}$ — number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$
Mean	np
Variance	$np(1-p)$
Entropy	$\frac{1}{2} \log_2 \left(2\pi e\, np(1-p)\right) + O\left(\frac{1}{n}\right)$ in shannons. For nats, use the natural log in the log.

Rule of large number

$$S = \sum_{i=1}^{\aleph} \varepsilon_i = \sum_{i=1}^{\aleph} \exp\left(N\phi_i\right) \approx \exp\left(N\phi_{max}\right) \qquad \aleph \propto N^p$$

Rule of large number

$$S = \sum_{i=1}^{\aleph} \varepsilon_i = \sum_{i=1}^{\aleph} \exp\left(N\phi_i\right) \approx \exp\left(N\phi_{max}\right) \qquad \aleph \propto N^p$$

Proof:

$$0 \leqslant \varepsilon_i \leqslant \varepsilon_{max}$$

$$\varepsilon_{max} \leqslant S \leqslant \aleph \varepsilon_{max}$$

$$\frac{\ln \varepsilon_{max}}{N} \leqslant \frac{\ln S}{N} \leqslant \frac{\ln \varepsilon_{max}}{N} + \frac{\ln \aleph}{N}$$

For $\aleph \propto N^p$

$$\frac{\ln \aleph}{N} = \frac{p \ln N}{N} \xrightarrow{N \to \infty} 0$$

$$\lim_{N \to \infty} \frac{\ln S}{N} = \frac{\ln \varepsilon_{max}}{N} = \phi_{max}$$

Method of steepest descent

$$I = \int \exp\left(N\phi\left(x\right)\right) dx$$

Method of steepest descent

$$I = \int \exp\left(N\phi\left(x\right)\right) dx$$

Hwo to extend this

$$I = \int \exp\left\{N\left[\phi\left(x_{max}\right) - \frac{1}{2}\left|\phi''\left(x_{max}\right)\right|\left(x - x_{max}\right)^2 + \cdots\right]\right\} dx$$

$$= e^{N\phi\left(x_{max}\right)} \int \exp\left[-\frac{N}{2}\left|\phi''\left(x_{max}\right)\right|\left(x - x_{max}\right)^2\right] dx$$

$$\cong \sqrt{\frac{2\pi}{N\left|\phi''\left(x_{max}\right)\right|}}\, e^{N\phi\left(x_{max}\right)}$$

Proof Stirling formula

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

Proof Stirling formula

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

$$\int_0^\infty x^N e^{-\alpha x} dx = \frac{N!}{\alpha^{N+1}}$$

$$\int_0^\infty e^{-\alpha x} dx = \frac{1}{\alpha}$$

Proof Stirling formula

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

$$\int_0^\infty x^N e^{-\alpha x} dx = \frac{N!}{\alpha^{N+1}}$$

$$\int_0^\infty e^{-\alpha x} dx = \frac{1}{\alpha}$$

$$\Gamma(N+1) = N! = \int_0^\infty x^N e^{-x} dx$$

$$= \int \exp\left[N \phi(x)\right] dx$$

$$\phi(x) = \ln x - \frac{x}{N}$$

Proof Stirling formula

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

$$\int_0^\infty x^N e^{-\alpha x} dx = \frac{N!}{\alpha^{N+1}}$$

$$\int_0^\infty e^{-\alpha x} dx = \frac{1}{\alpha}$$

$$\Gamma(N+1) = N! = \int_0^\infty x^N e^{-x} dx$$

$$= \int \exp\left[N \phi(x)\right] dx$$

$$\phi(x) = \ln x - \frac{x}{N}$$

$$0 = \left.\frac{d\phi}{dx}\right|_{x_{max}} = \frac{1}{x_{max}} - \frac{1}{N} = 0 \qquad x_{max} = N$$

$$\frac{d^2\phi}{d^2x} = -\frac{1}{x_{max}^2} = -\frac{1}{N^2}$$

Proof Stirling formula

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

$$\int_0^\infty x^N e^{-\alpha x} dx = \frac{N!}{\alpha^{N+1}}$$

$$\int_0^\infty e^{-\alpha x} dx = \frac{1}{\alpha}$$

$$\Gamma(N+1) = N! = \int_0^\infty x^N e^{-x} dx$$

$$= \int \exp\left[N\phi(x)\right] dx$$

$$\phi(x) = \ln x - \frac{x}{N}$$

$$0 = \left.\frac{d\phi}{dx}\right|_{x_{max}} = \frac{1}{x_{max}} - \frac{1}{N} = 0 \qquad x_{max} = N$$

$$\frac{d^2\phi}{d^2x} = -\frac{1}{x_{max}^2} = -\frac{1}{N^2}$$

$$N! = \int \exp\left[N\phi(x_{max}) - \frac{1}{2}\frac{1}{N^2}(x - x_{max})^2\right] dx$$

$$= \int \exp\left[N \ln N - N - \frac{1}{2N^2}(x - N)^2\right] dx$$

$$\approx N^N e^{-N} \sqrt{\frac{2\pi}{N\left|-1/N^2\right|}}$$

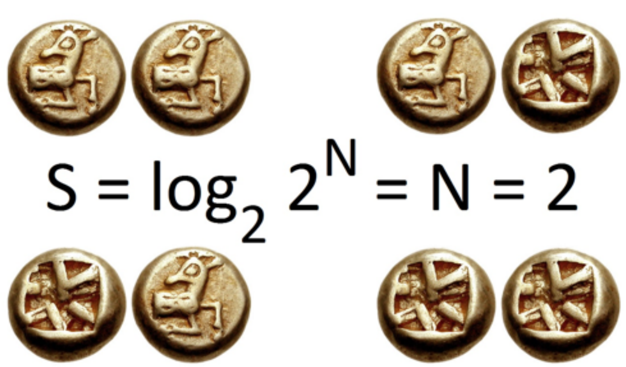
$$= N^N e^{-N} \sqrt{2\pi N}$$

$$\ln N! = N \ln N - N + \frac{1}{2} \ln 2\pi N + \mathcal{O}\left(\frac{1}{N}\right)$$

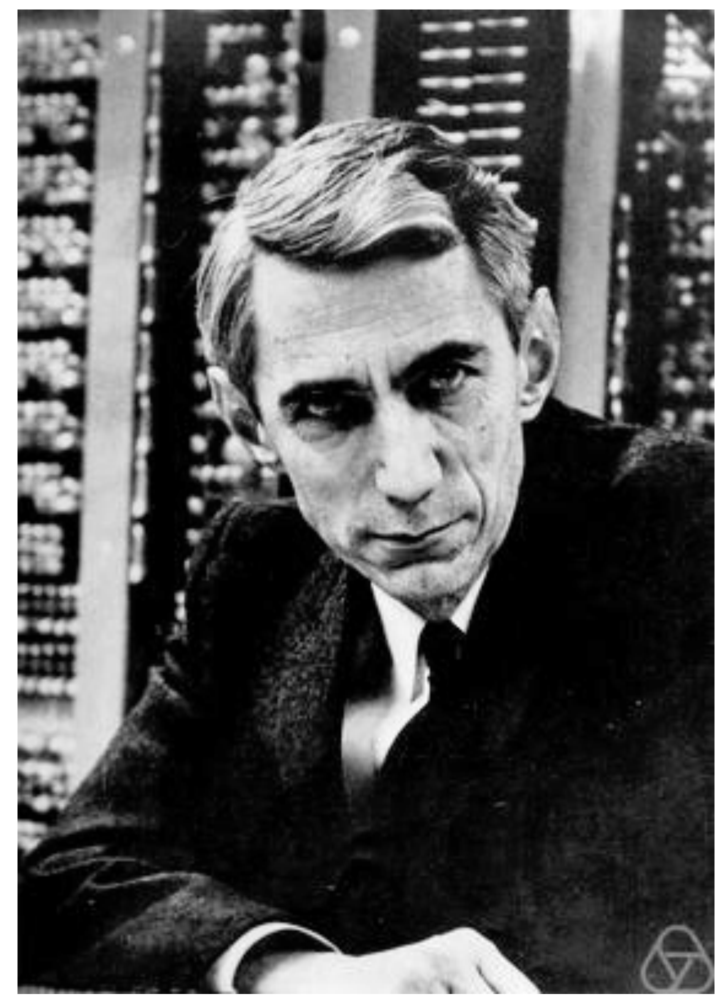
Entropy of information theory

information content of a probability distribution

In information theory, the entropy of a system represents the low boundary of bits to represent the information



$$S = \log_2 2^N = N = 2$$



Claude Shannon
(香农)

Entropy of information theory

Information entropy is typically measured in bits

$$\begin{array}{cc} A & B \\ 0 & 1 \end{array} \qquad \begin{array}{cccc} A & B & C & D \\ 00 & 01 & 10 & 11 \end{array}$$

$$\ln_2 2 = 1 \qquad \ln_2 4 = 2$$

A string with length N using M possible letter $N \ln_2 M$.

$g = M^N$ total # of possible string $\qquad \ln_2 g = \ln_2 M^N = N \ln_2 M$

- discrete set of outcomes $S = \{x_i\}, \quad i = 1, 2, \cdots, M$ occurring with probability P_i.
  A message from N indept outcomes of random varibales. There are M possibilities for each character in this message expect the message contain $N_i = N_{P_i}$ occur of each symbol.
  Number of typical messages corresponds to the number of ways of arranging the $\{N_i\}$ occurences of $\{x_i\}$

$$g = \frac{N!}{\prod_{i=1}^{M} N_i!}$$

Entropy of information theory

$$
\begin{aligned}
\ln_2 g &= \ln_2 \frac{N!}{\prod_{i=1}^{M} N_i!} \\
&= \ln_2 \frac{N!}{(NP_i)!} \\
&\approx \ln_2 \frac{N^N}{\prod_{i=1}^{M} (NP_i)^{NP_i}} \\
&= N \ln_2 N - \sum_{i=1}^{M} (NP_i) \ln_2 (NP_i) \\
&= N \ln_2 N - \sum_{i=1}^{M} NP_i \ln_2 N - \sum_{i=1}^{M} NP_i \ln_2 P_i \\
&= -N \sum_{i=1}^{M} P_i \ln_2 P_i
\end{aligned}
$$

$$H = -\sum_{i=1}^{M} P_i \log_2 P_i \quad \text{bits/symbol}$$

For uniform distribution $P_i = \frac{1}{M}$

$$
\begin{aligned}
&N \left(-\sum_{i=1}^{M} \frac{1}{M} \ln_2 \frac{1}{M} \right) \\
&= N \left(\sum_{i=1}^{M} \frac{1}{M} \right) \ln_2 M \\
&= N \ln_2 M
\end{aligned}
$$

If P_i is a δ-like function

$$N \left(-\sum_{i=1}^{M} P_i \ln P_i \right) = N\,(-1 \ln_2 1) = 0$$

Information entropy for any distribution

Shannon entropy

$$S = -\sum_{i=1}^{M} P(i) \ln P(i)$$

$$= -\langle \ln P(i) \rangle$$

Boltzman entropy

$$S = k_B \ln_2 [W]$$

W is the total number of possible microstates.

Information entropy and entropy of statistical mechanics

The entropy of statistical mechanics and the information entropy of information theory are basically the same thing.

$$S = -\sum_{i=1}^{M} p(i) \ln p(i)$$ (information theory)

$$S(E,\boldsymbol{x}) = -k_B \sum_i \frac{1}{\Omega} \ln \frac{1}{\Omega} = k_B \left(\sum_i \frac{1}{\Omega} \right) \ln \Omega = k_B \ln \Omega$$

$$S(E,\mathbf{x}) = k_B \ln \Omega(E,\mathbf{x})$$ (statistical mechanics)

E. T. Jaynes

Unbiased estimation

The principle of maximum entropy states that the probability distribution which best represents the current state of knowledge is the one with largest entropy

EX: If we know $\langle x \rangle = \mu$

$$\langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \Longrightarrow \langle x^2 \rangle = \sigma^2 + \mu^2$$

what is the $P(x)$?

Unbiased estimation

The principle of maximum entropy states that the probability distribution which best represents the current state of knowledge is the one with largest entropy

EX: If we know $\langle x \rangle = \mu$

$$\langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \Longrightarrow \langle x^2 \rangle = \sigma^2 + \mu^2$$

what is the $P(x)$?

maximize S w.r. t. constraints

Unbiased estimation

The principle of maximum entropy states that the probability distribution which best represents the current state of knowledge is the one with largest entropy

EX: If we know $\langle x \rangle = \mu$

$$\langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \Longrightarrow \langle x^2 \rangle = \sigma^2 + \mu^2$$

what is the $P(x)$?

maximize S w.r. t. constraints

$$S(\alpha, \beta, \gamma, \{P_i\}) = -\sum_i P(i) \ln P(i) - \alpha \left(\sum_i P_i - 1 \right) - \beta \left(\sum_i P_i x_i - \mu \right) - \gamma \left(\sum_i P_i x_i^2 - \left(\sigma^2 + \mu^2 \right) \right)$$

$$\frac{\delta S}{\delta P(i)} = -\ln P(i) - 1 - \alpha - \beta x_i - \gamma x_i^2 = 0$$

$$P(i) = e^{-\alpha - \beta x_i - \gamma x_i^2}$$

Principle of maximum entropy

Principle of maximum entropy

EX: If we know nothing

what is the $P(x)$?

Principle of maximum entropy

EX: If we know nothing

what is the $P(x)$?

EX: If we know $\langle x \rangle = \mu$

what is the $P(x)$?

Principle of maximum entropy

EX: If we know nothing

what is the $P(x)$?

EX: If we know $\langle x \rangle = \mu$

what is the $P(x)$?

EX: If we know $\langle F(x) \rangle = f$

what is the $P(x)$?

Principle of maximum entropy

EX: If we know nothing

what is the $P(x)$?

EX: If we know $\langle x \rangle = \mu$

what is the $P(x)$?

EX: If we know $\langle F(x) \rangle = f$

what is the $P(x)$?

A very powerful method to estimate a distribution !

Principle of maximum entropy

EX: If we know nothing

what is the $P(x)$?

EX: If we know $\langle x \rangle = \mu$

what is the $P(x)$?

EX: If we know $\langle F(x) \rangle = f$

what is the $P(x)$?

A very powerful method to estimate a distribution !

Macroscopic properties (Eq. of States) $\xleftarrow{SM}$ Microscopic Structure (Distribution) $\longleftarrow$ Most probable distribution (biggest number of microstate)

Summary

- Probability (Exponential/ Gaussian/Binomial)
- Rule of large number (Discrete/Steepest descent)
- Entropy (Information entropy/ Principle of maximum entropy)

Homework 2

PROBLEM 2: Find a the most probable distribution $P(x)$ such that,

$$\int P(x)\,dx = 1$$

$$\langle x \rangle = \int P(x)\,x\,dx = \mu$$

$$\left\langle x^2 \right\rangle - \langle x \rangle^2 = \int P(x)\,x^2\,dx - \left(\int P(x)\,x\,dx\right)^2 = \sigma^2$$

Here the most probable distribution is defined as the distribution $P(x)$ of which the entropy

$$S = -\int P(x) \ln P(x)\,dx$$

is maximized.

Compared the result with the problem 1 of chapter 1. What do you find?

PROBLEM 3: For different $N = 1, 10, 100, 1000$, numerically calculate the following functions,

$$\sum_{n=1}^{100} \exp\left[N/n^2\right]$$

Compared it with the function $\exp[N]$ (the $n = 1$ term, the largest term, in the above function), what do you find?

Homework

PROBLEM 4: The probability for observing a closed thermally equilibrated system with given energy E is in forms of

$$P(E) = C\Omega(E)\exp(-\beta E)$$

where C is the normalization constant and $\Omega(E)$ is the number of microstates with energy E. Both $\ln\Omega(E)$ and E are of the order of the particle number N. 1) Using the steepest descent method to verify that $P(E)$ is a very narrow distribution centered on the most probable value of E (denoted as E_0). 2) Use this expansion to estimated the probability for observing a spontaneous fluctuation in E of the order of $10^{-6}E_0$ for $1\,\mathrm{moles}$ of gas. (Hint: Expanding $\ln P(E)$ in powers of $\delta E = E - E_0$. Formular $S(E) = k_B \ln\Omega(E)$, $\partial E/\partial S = T$, $\partial E/\partial T = C_v$, $E_0 = 3Nk_BT/2$ and $C_v = 3Nk_B/2$ are needed in this calculation.)