Kardar 2.4

Probabilities with many random variables

$S_x = \{-\infty < x_1, x_2, \ldots, x_N < \infty\}$ set of outcomes

Joint Probability Distribution Function (PDF):
gives probability density of an outcome in vol. element $d^N x = \prod_{i=1}^{N} dx_i$ around pt. $\vec{x} = \{x_1, \ldots, x_N\}$

$$P_{\vec{x}}(S) = \int d^N x \, p(\vec{x}) = 1 \quad \text{normalization}$$

Iff the N random vars. are independent, then
$$p(\vec{x}) = \prod_{i=1}^{N} p_i(x_i).$$

Unconditional PDF: Describes probability of a subset of random variables, independent of the others.
$$p(x_1, \ldots, x_M) = \int \prod_{i=M+1}^{N} dx_i \; p(x_1, \ldots, x_N)$$

Example: Suppose a gas is composed of particles distributed in phase space with a probability distribution $p(\vec{x}, \vec{v})$.
If we are only interested in the distribution of velocities independent of positions, the unconditional PDF is
$$p_{\vec{v}}(\vec{v}) = \int d^3x \; p(\vec{x}, \vec{v})$$

Conditional PDF: Describes distribution of a subset of random variables for specified values of the others.

Example: The PDF of particle velocity at a given location $\vec{x}$, $\underbrace{p(\vec{v}\,|\,\vec{x})}_{\text{conditional PDF}}$, is given by:

$$p(\vec{v}\,|\,\vec{x}) = \frac{p(\vec{x},\vec{v})}{N}$$

, where the normalization is $N = \int d^3v\, p(\vec{x},\vec{v}) = p_{\vec{x}}(\vec{x})$, so that

$$\int d^3v\, p(\vec{v}\,|\,\vec{x}) = 1.$$

In general, $$\boxed{p(x_1,\ldots,x_M\,|\,x_{M+1},\ldots,x_N) = \frac{p(x_1,\ldots,x_N)}{p_{\vec{x}}(x_{M+1},\ldots,x_N)}}$$

This is related to Bayes' theorem:

$$p(A\,|\,B)\,p(B) = p(B\,|\,A)\,p(A)$$

Expectation values: $\langle F(\vec{x})\rangle = \int d^N x\, p(\vec{x})\, F(\vec{x})$

Joint characteristic function:

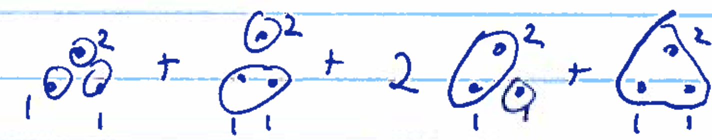
$$\tilde{p}(\vec{k}) = \left\langle \exp\left(-i\sum_{j=1}^{N} k_j x_j\right)\right\rangle$$

$$\langle x_1^{n_1}\cdots x_N^{n_N}\rangle = \left[\frac{\partial}{\partial(-ik_1)}\right]^{n_1}\cdots\left[\frac{\partial}{\partial(-ik_N)}\right]^{n_N}\tilde{p}(\vec{k})\Big|_{\vec{k}=\vec{0}}$$

$$\langle x_1^{n_1} * x_2^{n_2} * \cdots * x_N^{n_N}\rangle_c = \left[\frac{\partial}{\partial(-ik_1)}\right]^{n_1}\cdots\left[\frac{\partial}{\partial(-ik_N)}\right]^{n_N}\ln\tilde{p}(\vec{k})\Big|_{\vec{k}=\vec{0}}$$

Graphical relations between joint moments and joint cumulants require keeping track of labels on x_i:

$$\langle x_1 x_2 \rangle =$$  $$= \langle x_1 \rangle_c \langle x_2 \rangle_c + \langle x_1 * x_2 \rangle_c$$

$$\langle x_1^2 x_2 \rangle =$$ 

$$= \langle x_1 \rangle_c^2 \langle x_2 \rangle_c + \langle x_1^2 \rangle_c \langle x_2 \rangle_c + 2 \langle x_1 * x_2 \rangle_c \langle x_1 \rangle_c + \langle x_1^2 x_2 \rangle_c$$

Joint Gaussian Distribution

$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N \det C}} \exp\left[-\frac{1}{2}\sum_{mn} (C^{-1})_{mn} (x_m - \lambda_m)(x_n - \lambda_n)\right]$$

Inverse of some symmetric matrix C.

Joint characteristic function:

$$\tilde{p}(\vec{k}) = \exp\left[-i \sum_m k_m \lambda_m - \frac{1}{2}\sum_{mn} C_{mn} k_m k_n\right]$$

$$\Rightarrow \langle x_m \rangle_c = \lambda_m \quad , \quad \langle x_m * x_n \rangle_c = C_{mn}$$

If $\lambda_m = 0 \ \forall m$, all odd moments vanish, even moments obtained by summing over ways to group the relevant random variables in pairs:

$$\langle x_a x_b x_c x_d \rangle = C_{ab} C_{cd} + C_{ac} C_{bd} + C_{ad} C_{bc}$$

Central Limit Theorem

The sum $X = \sum_{i=1}^{N} x_i$ is a random variable, if x_i are random variables.

Suppose the x_i have a joint PDF $p(\vec{x})$.

The PDF for X is

$$p_X(x) = \int d^N x \, p(\vec{x}) \, \delta\left(x - \sum_{i=1}^{N} x_i\right)$$

$$= \int d^{N-1} x_i \, p\left(x_1, \ldots, x_{N-1}, x - x_1 - x_2 - \cdots - x_{N-1}\right)$$

Characteristic function for $p_X(x)$:

$$\tilde{p}_X(k) = \left\langle \exp\left(-ik \sum_{j=1}^{N} x_j\right)\right\rangle$$

$$= \tilde{p}\left(k_1 = k_2 = \cdots = k_N = k\right)$$

$$\ln \tilde{p}(k_1 = k_2 = \cdots = k_N = k) = -ik \sum_{i_1=1}^{N} \langle x_{i_1} \rangle_c$$

$$+ \frac{(-ik)^2}{2} \sum_{i_1, i_2 = 1}^{N} \langle x_{i_1} x_{i_2} \rangle_c + \cdots$$

$$\left.\frac{\partial}{\partial(-ik)}\right|_{k=0} \Rightarrow \langle X \rangle_c = \sum_{i_1} \langle x_{i_1} \rangle_c$$

$$\left.\frac{\partial^2}{\partial(-ik)^2}\right|_{k=0} \Rightarrow \langle X^2 \rangle_c = \sum_{i_1, i_2 = 1}^{N} \langle x_{i_1} x_{i_2} \rangle_c$$

If x_i are independent, $p(\vec{x}) = \prod_{i=1}^{N} p_i(x_i)$

$$\Rightarrow \tilde{p}_X(k) = \prod \tilde{p}_i(k)$$

$$\Rightarrow \langle X^n \rangle_c = \sum_{i=1}^{N} \langle x_i^n \rangle_c$$

If the PDF for each independent x_i is the same, then

$$\langle X^n \rangle_c = N \langle x^n \rangle_c \quad \text{, where } \langle x_i^n \rangle_c \equiv \langle x^n \rangle_c \ \forall i.$$

$$\langle X \rangle \propto N, \quad \sqrt{\langle (X - \langle X \rangle)^2 \rangle} \propto \sqrt{N}.$$

Consider the random variable $y \equiv \frac{1}{\sqrt{N}} \left(X - N \langle x \rangle_c \right)$.

$$\langle y \rangle = 0, \quad \langle y^n \rangle_c \propto N^{1 - n/2}$$

As $N \to \infty$, only $\langle y^2 \rangle_c \propto N^0$ survives,

$$\langle y^n \rangle_c \xrightarrow{N \to \infty} 0 \quad \text{for } n > 2.$$

The only distribution with only first and second cumulants is the Gaussian distribution.

$$\Rightarrow \boxed{\lim_{N \to \infty} p\left(y = \frac{\sum_{i=1}^{N} x_i - N \langle x \rangle_c}{\sqrt{N}} \right) = \frac{1}{\sqrt{2\pi \langle x^2 \rangle_c}} \exp\left(-\frac{y^2}{2 \langle x^2 \rangle_c} \right)}$$

This is the <u>Central Limit Theorem</u>.

Even if the random variables are not independent,

$$\sum_{i_1,\dots,i_m=1}^{N} \langle x_{i_1}\cdots x_{i_m}\rangle_c \ll O(N^{m/2}),$$

so it is still the case that only the (first and) second cumulants of $y = \frac{\sum_{i=1}^{N}(x_i - N\langle x_i\rangle_c)}{\sqrt{N}}$ survive

$\Rightarrow p(y) \xrightarrow{N\to\infty}$ Gaussian as before.

Hence, the Central Limit Theorem is more general than the result when the x_i are mutually independent.

Kardar 2.6

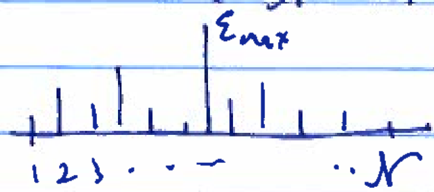
Rules for large numbers

Thermodynamic limit: # microscopic degrees of freedom $N \to \infty$.

Intensive quantities: $O(N^0)$ (e.g. T, P)
Extensive quantities: $O(N)$ (e.g. E, S)
Exponential dependence: $O(\exp(N\phi))$ (e.g. # microstates)

Sums of exponential quantities are dominated by the largest term:
Consider $S = \sum_{i=1}^{\mathcal{N}} \mathcal{E}_i$ with $0 \le \mathcal{E}_i \sim O(\exp(N\phi_i))$,

terms $\mathcal{N} \propto N^p$ for some positive power p.



$$\mathcal{E}_{max} \le S \le \mathcal{N}\,\mathcal{E}_{max}.$$

Can construct an <u>intensive</u> quantity: $\frac{1}{N}\ln S$

$$\frac{\ln \mathcal{E}_{max}}{N} \le \frac{\ln S}{N} \le \frac{\ln \mathcal{E}_{max}}{N} + \cancel{\frac{p\ln N}{N}}$$

vanishes as $N\to\infty$.

$$\Rightarrow \lim_{N\to\infty} \frac{\ln S}{N} = \frac{\ln \mathcal{E}_{max}}{N} = \phi_{max}.$$

Continuum generalization: <u>Saddle point Integration</u>

Consider $J = \int dx \exp\left(N\phi(x)\right)$ as $N\to\infty$.

Expand about x_{max} that maximizes $\phi(x)$:

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

$\phi''(x_{max}) \le 0$

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

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

Stirling's Approximation for $N!$ as $N \gg 1$:

$$\Gamma(N+1) = \int_0^\infty dx\, x^N e^{-x} = N! \text{ for } N \text{ integer.}$$

Write $\Gamma(N+1) = \int_0^\infty dx \exp\left(N\phi(x)\right)$ with $\phi(x) = \ln x - \frac{x}{N}$

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

$$0 = \phi'(x_{max}) = \frac{1}{x_{max}} - \frac{1}{N} \rightarrow x_{max} = N$$

$$\phi(x_{max}) = \ln N - 1$$

$$\phi''(x_{max}) = -\frac{1}{x_{max}^2} = -\frac{1}{N^2}$$

$$\Gamma(N+1) \approx \int_0^\infty dx \exp\left[N \ln N - N - \frac{N}{2N^2}(x-N)^2\right]$$

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

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

Stirling's Formula.

Kardar 2.7

Information and Entropy

Suppose a random variable has a discrete set of outcomes $S = \{x_i\}$, with probabilities $\{p_i\}$, $i = 1, \ldots, M$.

Suppose we compose a message from N independent outcomes, each w/ M possibilities.
The number of bits (0 or 1) necessary to convey the message precisely is $N \log_2 M$, because there are $2^{(N \log_2 M)} = M^N$ possible messages.

If drawn from the random distribution, for large N there are roughly $N_i = N p_i$ occurrences of x_i in the message.

The number of typical messages is the number of ways of arranging the $\{N_i\}$ occurrences of $\{x_i\}$ in the N outcomes:

$$g = \frac{N!}{N_1! N_2! \cdots N_M!}, \text{ where } \sum_{i=1}^{M} N_i = N.$$

The number of bits required to send one of the g possible typical messages is

$$\boxed{\log_2 g} \approx \log_2 N! - \sum_{i=1}^{M} \log_2 N_i!$$

$$\overset{\text{Stirling}}{\approx} N \log_2 N - \sum_{i=1}^{M} N_i \log_2 N_i$$

$$= -N \sum_i \frac{N_i}{N} \log_2 \frac{N_i}{N} \quad \boxed{\approx -N \sum_i p_i \log_2 p_i}$$

For a uniform distribution $p_i = \frac{1}{M}$.

$$\log_2 g_{uniform} \approx -N \sum_{i=1}^{M} \frac{1}{M} \log_2 \frac{1}{M}$$

$$= +N \log_2 M$$

Other distributions give $\log_2 g < N \log_2 M$.

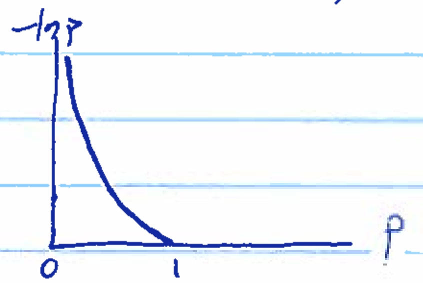
If only one outcome is almost certainly possible, say x_1, then $p_1 = 1$, $p_{i \neq 1} = 0$.

Then, $\log_2 g \approx -N \cdot 1 \cdot 1 \log_2 1 = 0$.

Entropy: The entropy associated to a probability distribution $\{p_i\}$ is defined as $(i = 1, \ldots, M)$

$$\boxed{S \equiv -\sum_{i=1}^{M} p_i \ln p_i = -\langle \ln p_i \rangle}$$

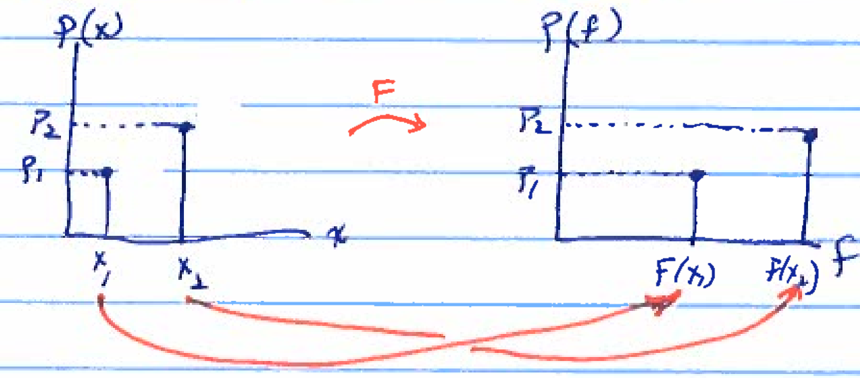
You can think of $-\ln p_i$ as a measure of the surprise at observing outcome x_i.



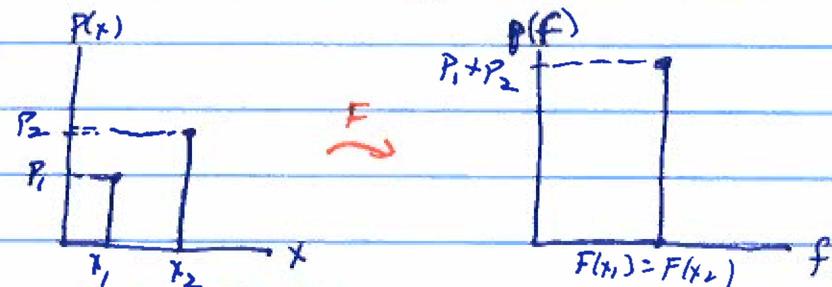
The entropy is the average surprise for a given distribution.

Entropy does not depend on the values of the random variable x, only the probabilities associated with possible outcomes $\{x_i\}$.

The probability distribution associated with a random variable $f = F(x)$ has the same entropy as the probability distribution of x.



A many-to-one map decreases the entropy:



$$\boxed{\Delta S(x_1, x_2 \to f)} = (P_1 \ln P_1 + P_2 \ln P_2) - (P_1+P_2) \ln (P_1+P_2)$$

$$= P_1 \ln \frac{P_1}{P_1+P_2} + P_2 \ln \frac{P_2}{P_1+P_2} \quad \boxed{< 0}$$

Hence, a coarse graining that does not distinguish between certain subsets of outcomes reduces the entropy.