Differential entropy and Shannon's entropy: information theory lecture 8

COMSM0075 Information Processing and Brain

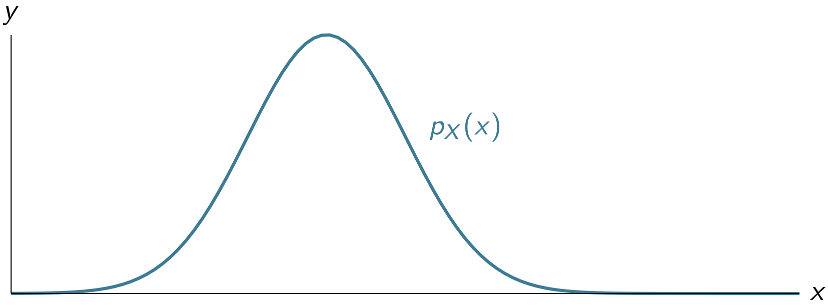
`comsm0075.github.io`

October 2020

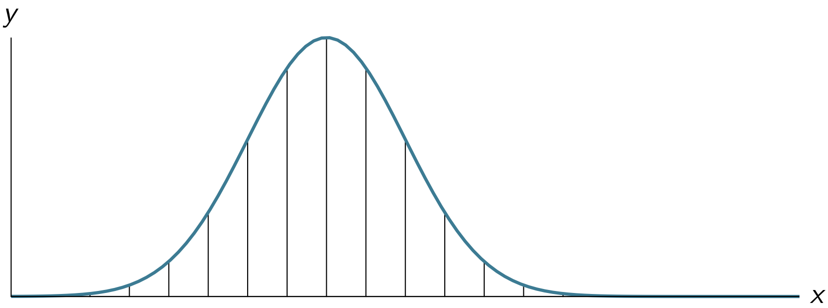
Differential entropy

$$h(X) = -\int dx p_X(x) \log_2 p_X(x)$$

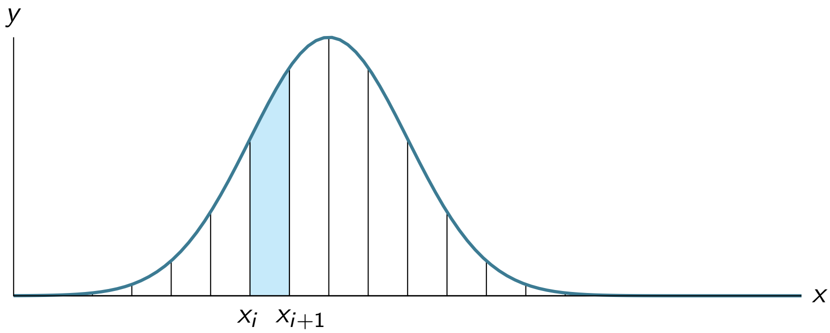
A continuous distribution



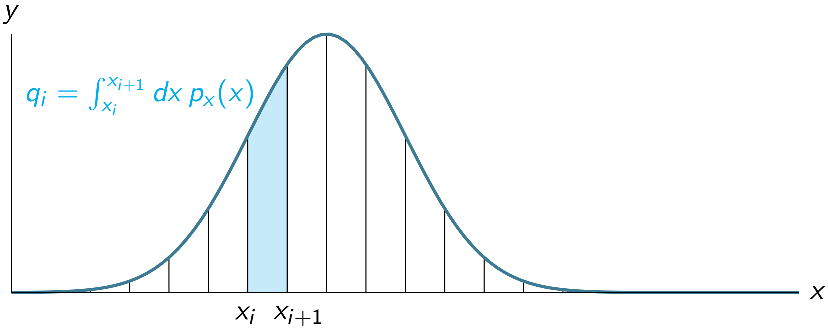
A continuous distribution



Discrete from continuous



Discrete from continuous



A discrete random variable

Let $\delta = x_{i+1} - x_i$.

Consider the discrete random variable X^δ whose outcomes are the $\{x_1, x_2, \ldots\}$ and whose probabilities are given by

$$p_{X^\delta}(x_i) = q_i$$

Shannon's entropy

$$p_{X^\delta}(x_i) = q_i$$

so

$$H(X^\delta) = -\sum_i q_i \log_2 q_i$$

Get rid of the integrals

$$q_i = \int_{x_i}^{x_{i+1}} dx\, p_x(x)$$

The plot obviously is to get rid of the integral here, morally:

$$q_i = \int_{x_i}^{x_{i+1}} dx\, p_X(x) \approx p_X(\bar{x}_i)\delta$$

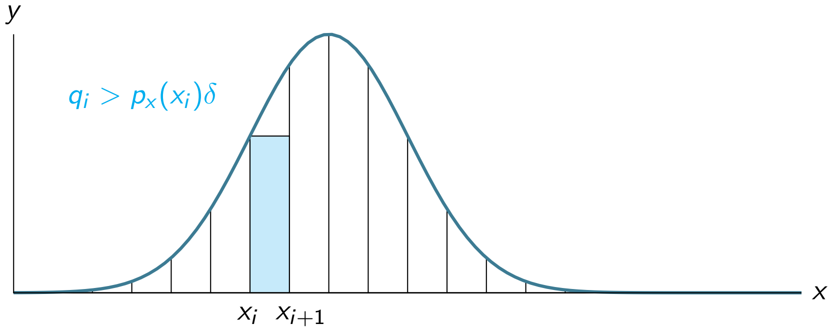
where $\bar{x}_i$ is any point in $[x_i, x_{i+1})$ and the approximation gets better as δ gets smaller.

Mean Value Theorem

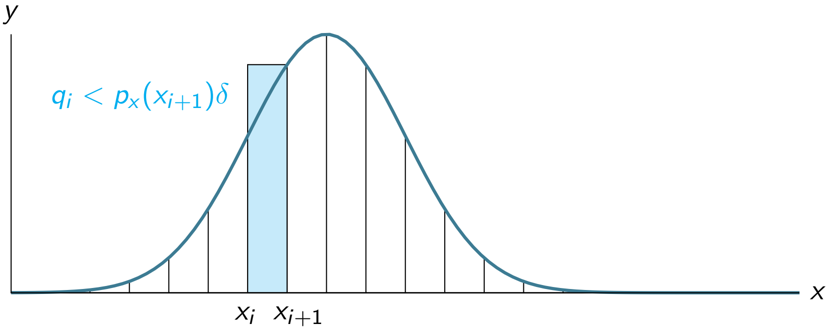
$$q_i = \int_{x_i}^{x_{i+1}} dx \, p_X(x) \approx p_X(\bar{x}_i)\delta$$

Rather than worry about how good the approximation is, we use an elegant alternative approach based on the mean value theorem.

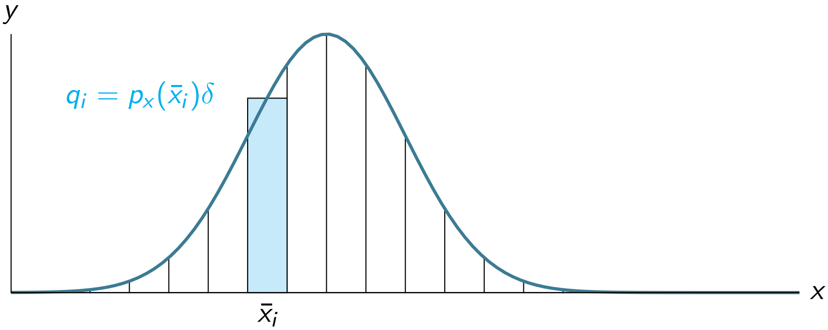
Mean Value Theorem



Mean Value Theorem



Mean Value Theorem



Mean Value Theorem

The **Mean Value Theorem** tells us that assuming $p_X(x)$ is continuous we know we can always pick a value of $\bar{x}_i \in [x_i, x_{i+1}]$ such that

$$p(\bar{x}_i)\delta = q_i = \int_{x_i}^{x_{i+1}} p_x(x)dx$$

exactly.

Integrals are gone

$$H(X^\delta) = -\sum_i q_i \log_2 q_i = -\sum_i p_X(\bar{x}_i)\delta \log_2 p_X(\bar{x}_i)\delta$$

or, expanding out the log and using

$$\sum_i p_X(\bar{x}_i)\delta = \int dx\, p_X(x) = 1$$

we get

$$H(X^\delta) = -\sum_i p_X(\bar{x}_i)\delta \log_2 p_X(\bar{x}_i) - \log_2 \delta$$

Use the Riemann approximation

$$\sum_i f(\bar{x}_i)\delta \to \int f(x)dx$$

as $\delta \to 0$.

Finally

$$H(X^{\delta}) + \log_2 \delta = -\sum_i p_X(\bar{x}_i)\delta \log_2 p_X(\bar{x}_i) \to h(X)$$

as $\delta \to 0$.

Shannon's entropy and differential entropy

$$H(X^\delta) + \log_2 \delta \approx h(X)$$

for small enough δ.

Shannon's entropy and differential entropy

$$H(X^\delta) + \log_2 \delta \approx h(X)$$

for small enough δ.