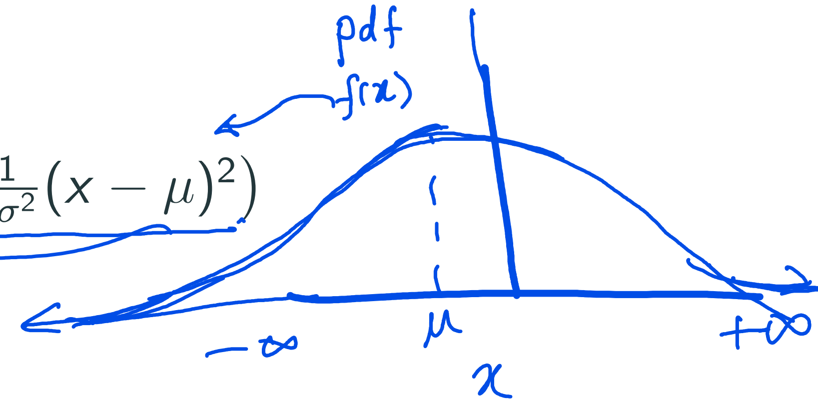
Normal (Gaussian) Distribution

- It is a popular continuous distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$



  - μ is the mean and σ^2 is the variance

- Exercise: Verify the mean and variance. For e.g.

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx \stackrel{?}{=} \mu$$

$$= \int_{-\infty}^{\infty}(x-\mu)\,N(x|\mu,\sigma^2)\,dx + \int_{-\infty}^{\infty}\mu\,N(x|\mu,\sigma^2)\,dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}(\dots)\Big|_{-\infty}^{+\infty} + \dots \qquad = 1$$

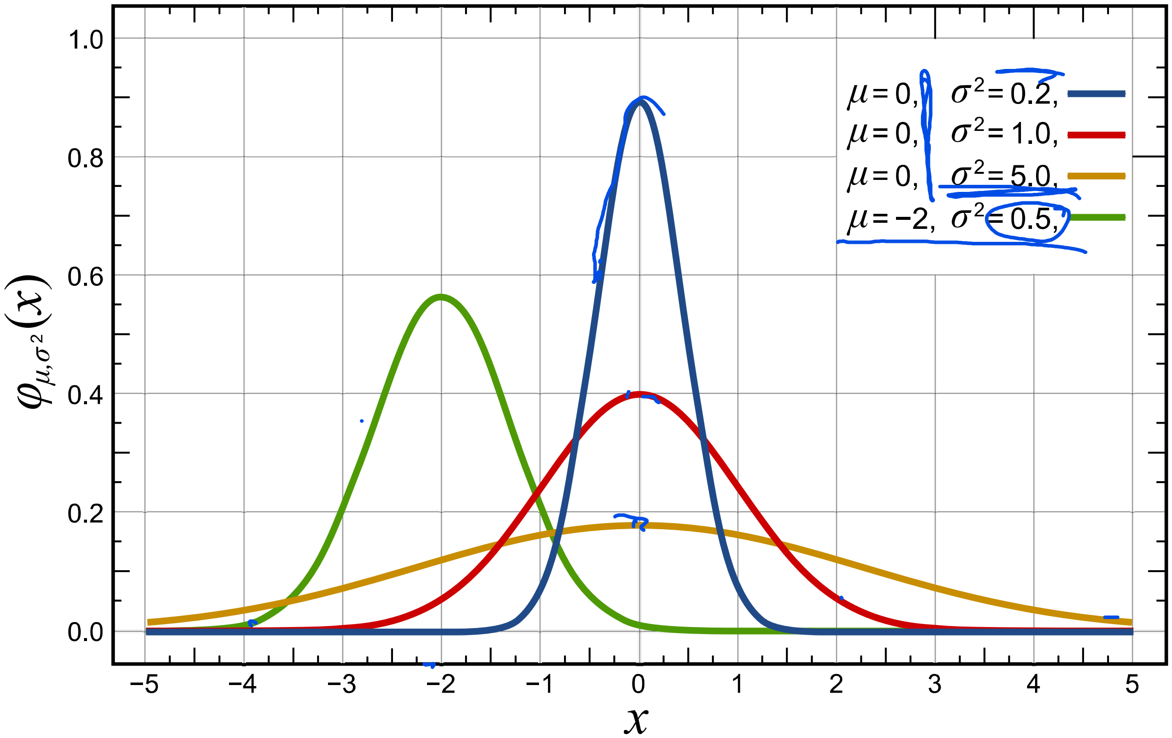
$E[X-\underbrace{E(X)}_{=\mu}]^2$

$\int_{-\infty}^{\infty}(x-\mu)^2 N(x|\mu,\sigma^2)\,dx$

$\vdots$

σ^2

1-D Gaussian distribution



Normal (Gaussian) Distribution

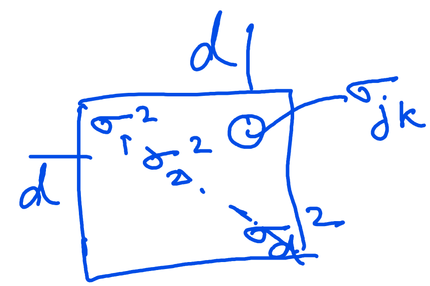
- It is a popular continuous distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

  - μ is the mean and σ^2 is the variance

- Exercise: Verify the mean and variance. For e.g.

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx \stackrel{?}{=} \mu$$



- Multivariate Gaussian (d-dim) $\mu \in R^d$ $\Sigma \in \mathbb{R}^{d \times d}$

$x \in \mathbb{R}^d$

$$f(x|\mu,\Sigma) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

$\rightarrow$ determinant

  - x is now a vector, μ is the mean vector and Σ is the co-variance matrix

2-D Gaussian distribution

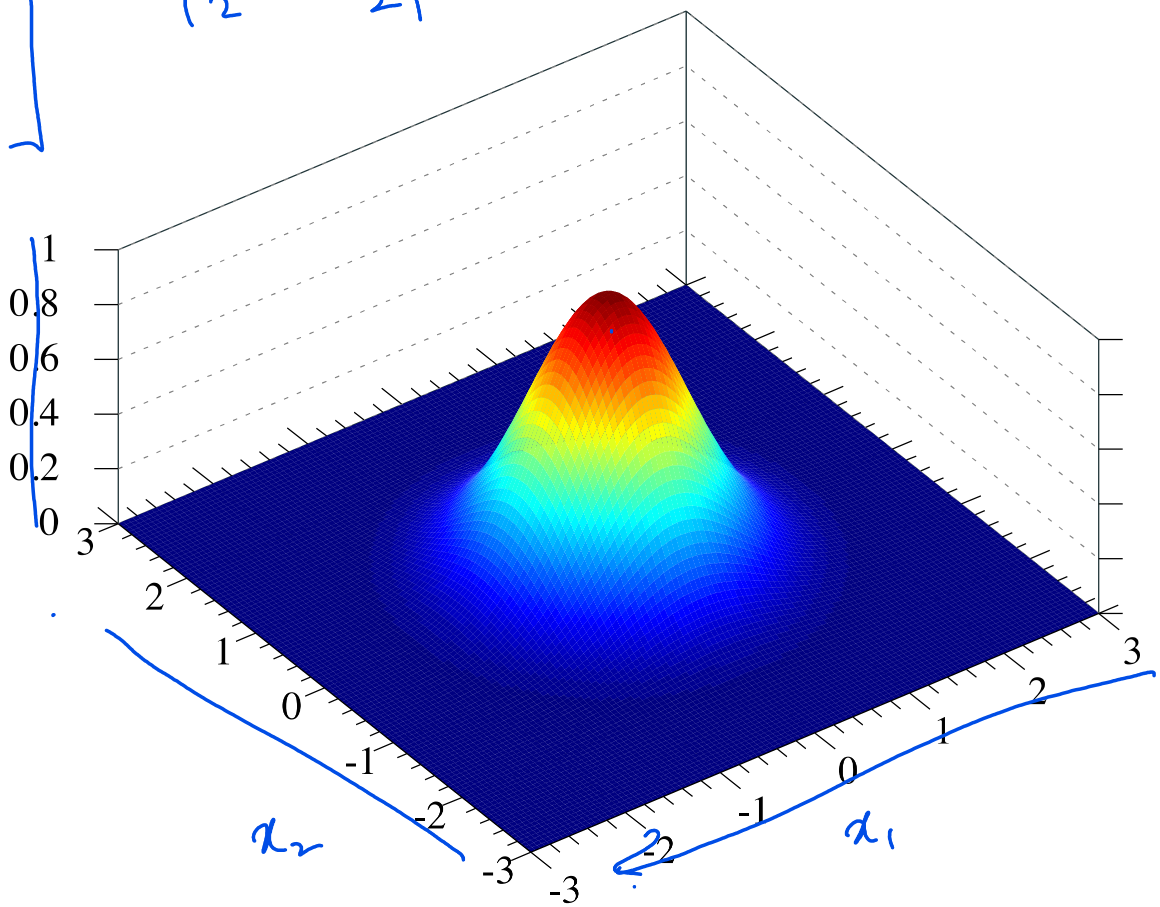
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

symmetric

$$\sigma_{12} = \sigma_{21}$$

$$\frac{1}{(\sqrt{2\pi})|\Sigma|^{1/2}} \exp\left(\cdot - \right)$$

$d = 2$



Properties of Normal Distribution

- All marginals of a Gaussian are again Gaussian
- Any conditional of a Gaussian is Gaussian
- The product of two Gaussians is again Gaussian
- Even the sum of two independent Gaussian r.v.'s is a Gaussian

$p(x_1, x_2) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \Sigma\right)$

$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$

$p(x_1) = \mathcal{N}(x_1; u_1, \sigma_1^2)$

$p(x_1 | x_2 = c) = \mathcal{N}(x_1;$

$\mu_{x_1|c} = g(c, u_2, \sigma_{12})$

$\sigma_{x_1|c} = h(\sigma_1, \sigma_2, \sigma_{12})$

Multinomial distribution

$X \sim \text{Multi}(x;\ \underbrace{p_1, p_2, \dots p_K}_{\text{parameters.}})$ $\quad X = \{1, 2, \dots K\}$

$P(X = v) = p_v \qquad \sum_v P(X=v) = 1 \Rightarrow \sum_v p_v = 1$

Example: rolling of a dice.

$X = \{1, 2, \dots 6\} \qquad X \sim \text{Mult}(x;\ 0.1, 0.2, 0.1, 0.5, 0.1)$

Exponential family distributions

$*$ Not in syllabus.

Many of the standard distributions belong to this family

- Bernoulli, binomial/multinomial, Poisson, Normal (Gaussian), Beta/Dirichlet ...
- Share many important properties - e.g. They have a conjugate prior.

$$p(x; \eta \in \mathbb{R}^p) = h(x) \exp\left(\eta^T T(x) - A(\eta)\right)$$

Samples of a Random Variable

Let X be a R.V with probability function $p(x) = f(x; \mathbf{w})$ parameter.

Samples of X are set of values $\{x^1, \ldots, x^N\}$ assigned to X based on $p(x)$.

Examples: $X \sim \text{Bernoulli}(x;\ q = 0.9)$

$N = 10$ 0, 0, 1, 1, 1, 0, 1, 1, 1, 0

$N = 10^6$ → 1, 1, 0, 0, 1, 1, 1, 1, 1, 1

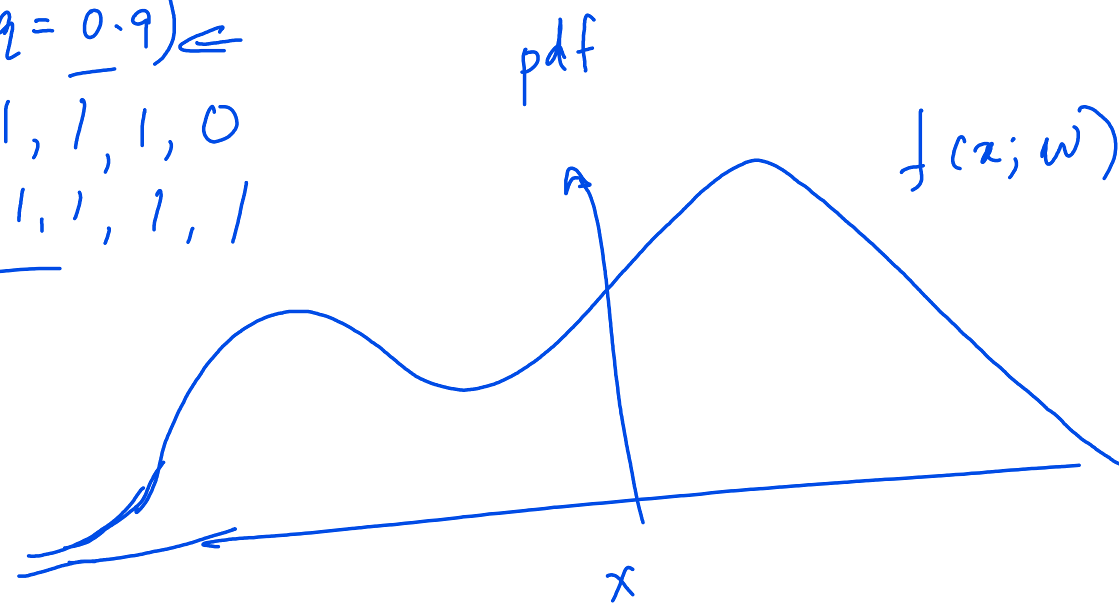
D^1

D^2

.

.

D^{10}



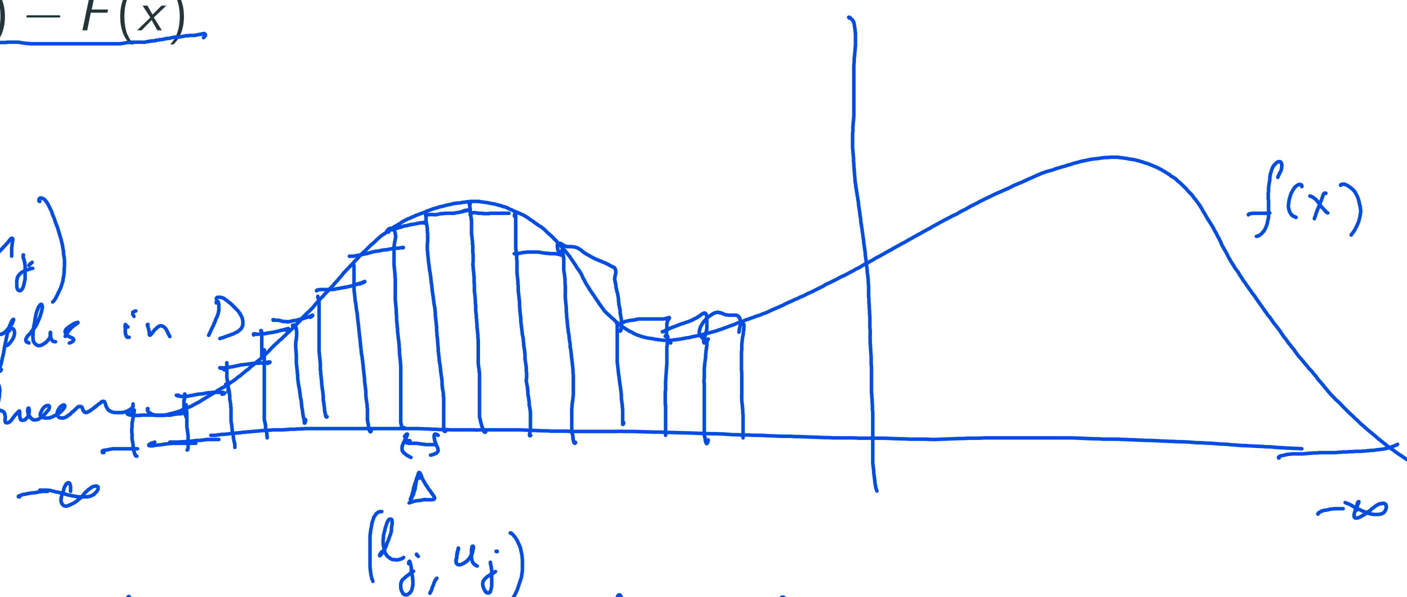
Consistent samples

As $N \to \infty$, the fraction of times in the sample that we encounter a sample in an interval $[x, x + \Delta)$ would be proportional to the true probability of that interval in $p(x)$ that is, $F(x + \Delta) - F(x)$

D

For every interval (l_j, u_j)

find $n_j \equiv$ # of samples in D whose value is between (l_j, u_j)

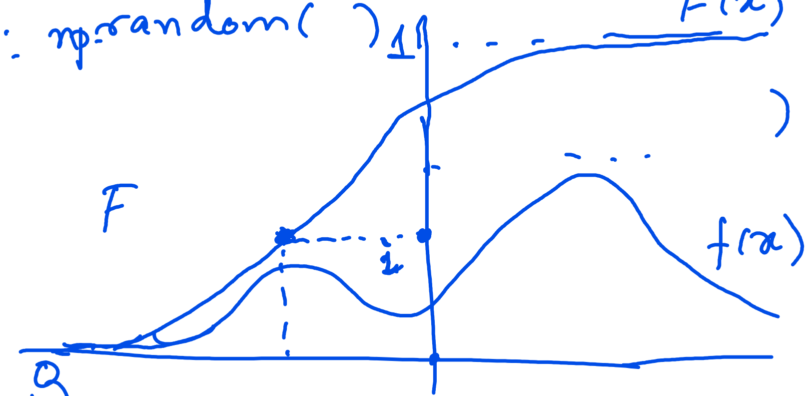


Plot $\left(\frac{n_j}{N}\right)$ against x and compare with true $f(x)$

How to draw samples?

$f(x; w)$

How do we draw N samples $x^1, \ldots, x^N$ from the distribution? Assume we can sample a u from a uniform distribution $U(0, 1)$ Eg: np.random( )



Let $F(x)$ be cumulative distribution of $p(x) = f(x; w)$

$u_i = 0.3$

Find the x at which $F(x) = u_i = 0.3$.

$\Rightarrow x = F^{-1}(u_i)$

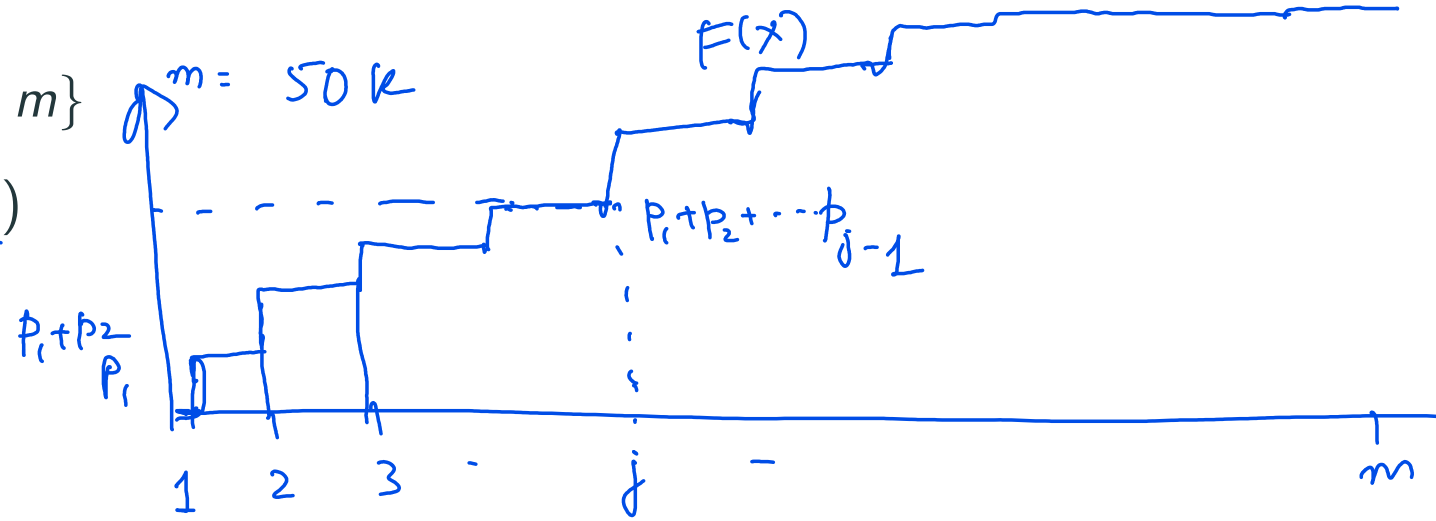
For $i = 1 \ldots N$

1. Sample $u_i \sim U(0, 1)$
2. Find $x^i = F^{-1}(u_i)$

Basics: Sampling from multinomial distributions

x is discrete, $x \in \{1, \ldots m\}$

$p(x) \sim Mult(p_1, \ldots, p_m)$



for $i = 1$ to N

$u_i \sim U(0, 1)$

find the r s.t u_i is between $F(r)$ and $F(r+1)$

Output r.

Parameter Estimation

Given samples $D = \{x^1, \dots, X^N\}$, form of the distribution $p(x) = f(x; \mathbf{w})$ estimate values of the parameters w.

i.i.d assumption: each instance is independently and identically distributed.

Two methods:

1. Maximum likelihood estimation
2. Bayesian estimation

Maximum Likelihood Estimation (MLE)

- Find the value $\mathbf{w}$ for which the probability (likelihood) of the data is maximized.
- Likelihood of data $L(w; D)$
- Maximizing log-likelihood of data is equivalent to maximizing likelihood.