Fundamentals of Machine Learning

UNIVARIATE AND MULTIVARIATE GAUSSIAN

Amit K Roy-Chowdhury

Acknowledgments: Adapted from slides at https://probml.github.io/pml-book/teaching1.html by Prof. Saw Shier Nee

Univariate Gaussian Distribution

The most widely used distribution of real-valued random variables is the Gaussian distribution, also called the normal distribution.

The following shows the cumulative distribution function of a continuous random variable Y.

$$P(y) \triangleq \Pr(Y \le y)$$

The following shows the probability of Y between a and b is given as following.

$$\Pr(a < Y \le b) = P(b) - P(a)$$

Cdf of Gaussian is defined as

$$\Phi(y; \mu, \sigma^2) \triangleq \int_{-\infty}^{y} \mathcal{N}(z|\mu, \sigma^2)dz = \frac{1}{2}[1 + \text{erf}(z/\sqrt{2})] \quad , \quad z = (y - \mu)/\sigma, \quad \text{erf}(u) \triangleq \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt$$

mean variance

Gaussian Distribution

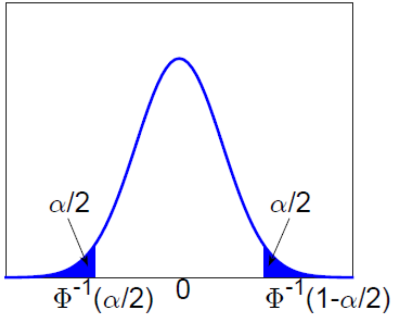
The most widely used distribution of real-valued random variables is the Gaussian distribution, also called the normal distribution.

Probability distribution function of Gaussian is

$$\mathcal{N}(y|\mu,\sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$$

mean variance

Normalization – ensure area under curve = 1

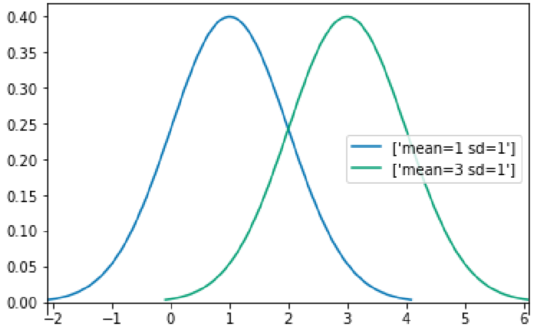


Inverse cdf

Gaussian Distribution

Mean of pdf of Gaussian is = 0

$$\mathbb{E}[Y] \triangleq \int_{\mathcal{Y}} y\, p(y) dy$$

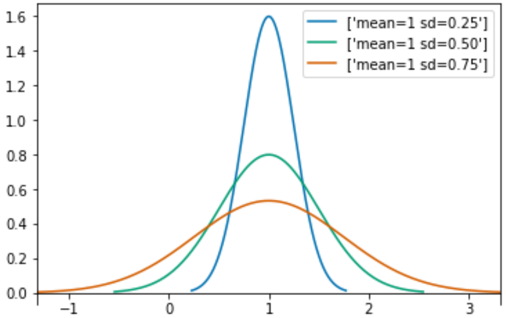


Variance of pdf of Gaussian is = Spread

$$\mathbb{V}[Y] \triangleq \mathbb{E}\left[(Y-\mu)^2\right] = \int (y-\mu)^2 p(y) dy$$

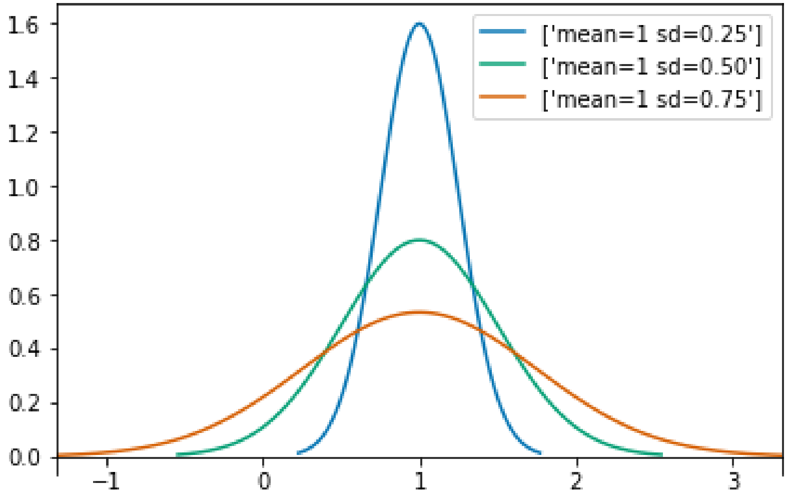
$$= \int y^2 p(y) dy + \mu^2 \int p(y) dy - 2\mu \int y p(y) dy = \mathbb{E}\left[Y^2\right] - \mu^2$$

$$\mathbb{E}\left[Y^2\right] = \sigma^2 + \mu^2$$



Dirac delta function

When variance of Gaussian goes to zero, the distribution becomes narrower.



$$\lim_{\sigma \to 0} \mathcal{N}(y|\mu, \sigma^2) \to \delta(y - \mu)$$

where δ is the **Dirac delta function**, defined by

$$\delta(x) = \begin{cases} +\infty & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases}$$

where

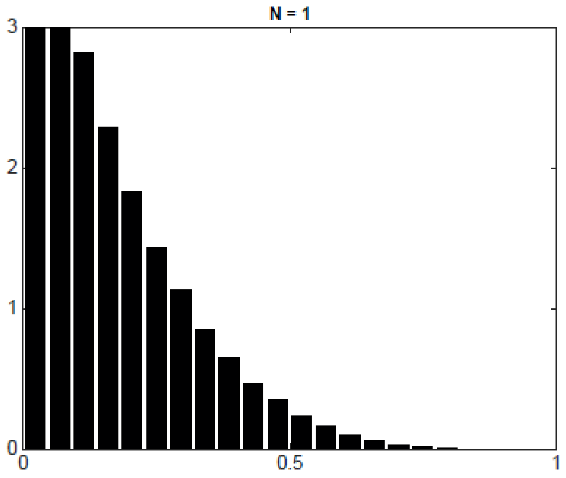
$$\int_{-\infty}^{\infty} \delta(x) dx = 1$$

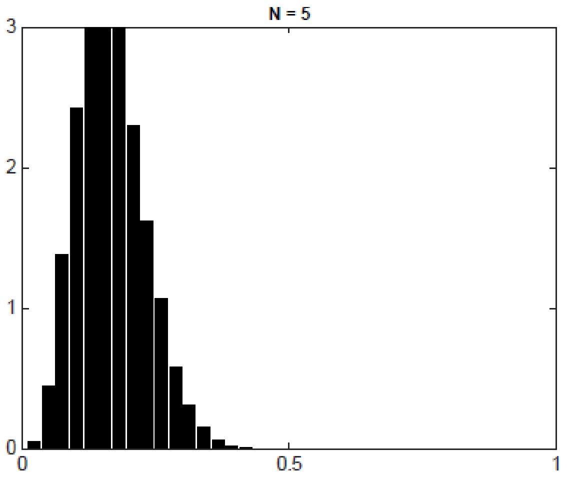
Sifting Property

$$\int_{-\infty}^{\infty} f(y)\delta(x - y) dy = f(x)$$

Central Limit Theorem

When independent random variables are summed up, the distribution converges to a normal distribution even if the original distribution themselves are not normally distributed.

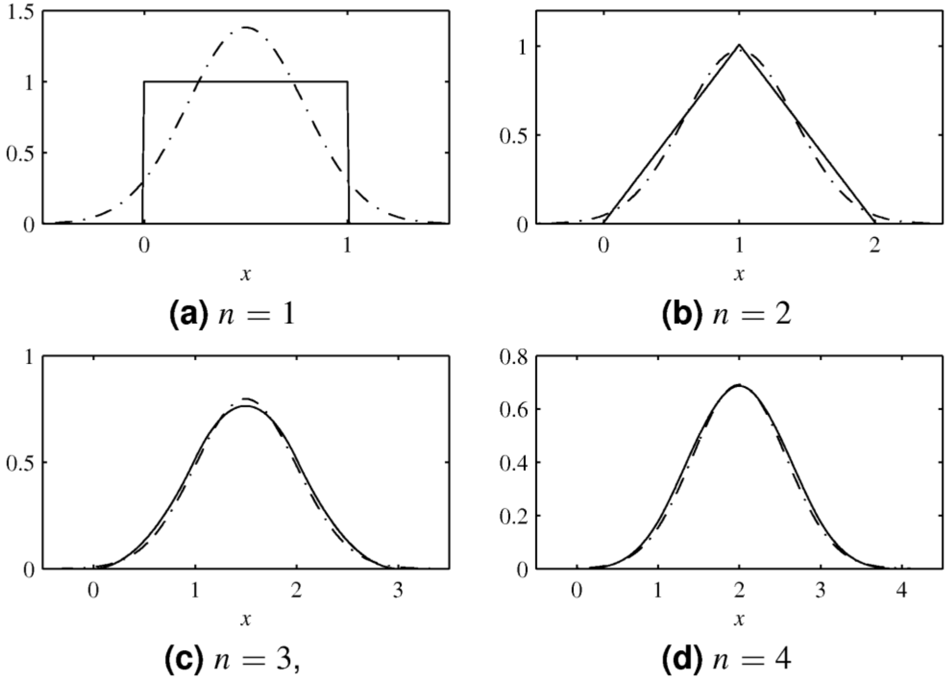




PDF of sum of random variables

When X and Y are independent random variables, the PDF of $W = X + Y$ is

$$f_W(w) = \int_{-\infty}^{\infty} f_X(w-y)\, f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x)\, f_Y(w-x)\, dx.$$



The PDF of W_n, the sum of n uniform $(0, 1)$ random variables, and the corresponding central limit theorem approximation for $n = 1, 2, 3, 4$. The solid — line denotes the PDF $f_{W_n}(w)$, while the $- \cdot -$ line denotes the Gaussian approximation.

PDF of sum of random variables

The sum of n independent Gaussian random variables $W = X_1 + \cdots + X_n$ is a Gaussian random variable.

Multivariate - Gaussian

The most widely used joint probability distribution for continuous random variable is the multivariate Gaussian or multivariate normal (MVN).

MVN density is

$$\mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right]$$

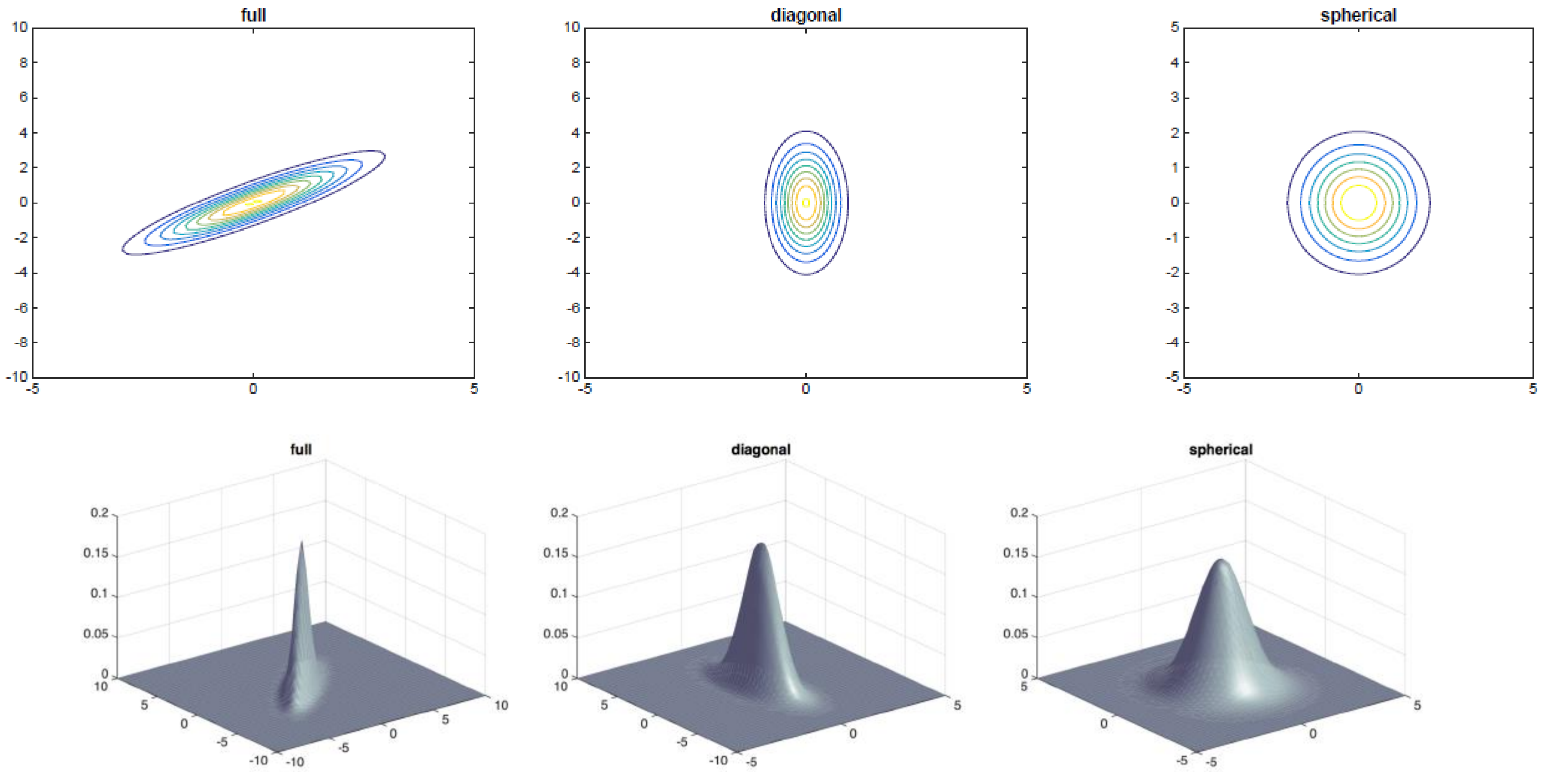
Mean vector

Covariance matrix

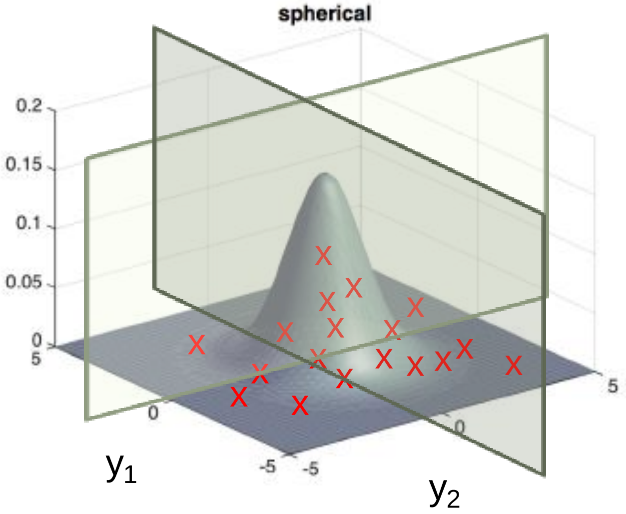
$$\mathrm{Cov}\left[\boldsymbol{y}\right] \triangleq \mathbb{E}\left[(\boldsymbol{y}-\mathbb{E}\left[\boldsymbol{y}\right])(\boldsymbol{y}-\mathbb{E}\left[\boldsymbol{y}\right])^{\mathsf{T}}\right]$$

$$\mathbb{E}\left[\boldsymbol{y}\boldsymbol{y}^{\mathsf{T}}\right] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}}$$

Multivariate - Gaussian



Conditional Gaussian



Posterior Conditional Formula

$$
\begin{aligned}
p(y_1|y_2) &= \mathcal{N}(y_1|\mu_{1|2}, \Sigma_{1|2}) \\
\mu_{1|2} &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2) \\
&= \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(y_2 - \mu_2) \\
&= \Sigma_{1|2}\left(\Lambda_{11}\mu_1 - \Lambda_{12}(y_2 - \mu_2)\right) \\
\Sigma_{1|2} &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Lambda_{11}^{-1}
\end{aligned}
$$