Gaussian Distribution

Tuan Nguyen

Ngày 28 tháng 7 năm 2021

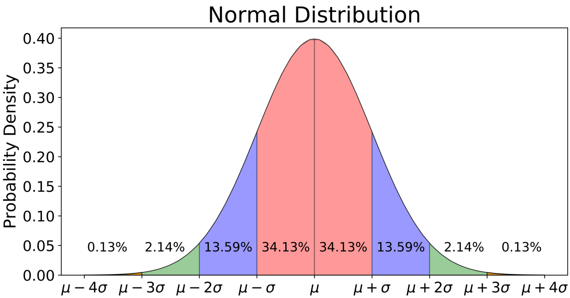
Overview

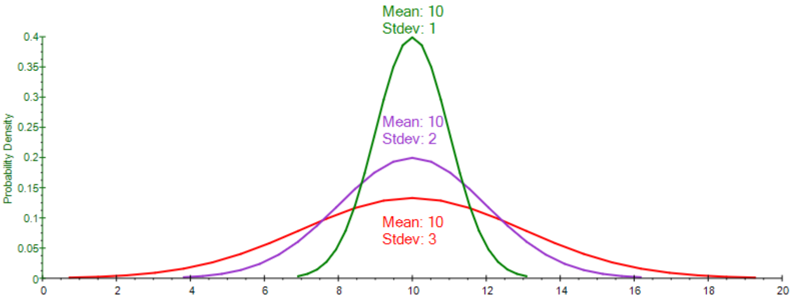
Gaussian distribution

Gaussian distribution (normal distribution) is the most well-studied probability distribution for continuous-valued random variables.

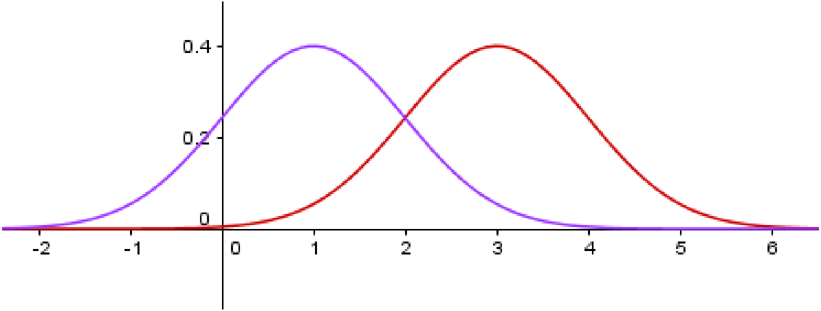
Univariate Gaussian

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





Hình 1: Same mean and different standard deviation



Hình 2: Same standard deviation and different mean

Gaussian distribution (cont.)

- Normalization

$$\int_{-\infty}^{\infty} p(x|\mu, \sigma^2) = 1$$

- Mean

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} xp(x|\mu, \sigma^2) = \mu$$

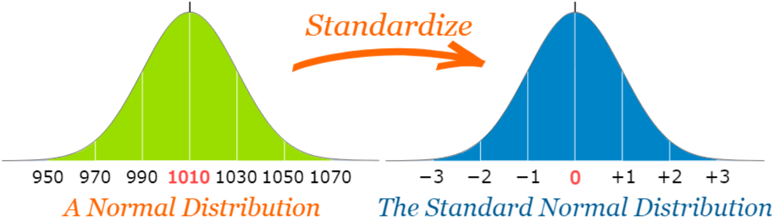
- Variance

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x|\mu, \sigma^2) = \sigma^2$$

Gaussian distribution (cont.)

Standardize the normal distribution to standard normal distribution.

$$X \sim \mathcal{N}(\mu, \sigma) \Rightarrow Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$



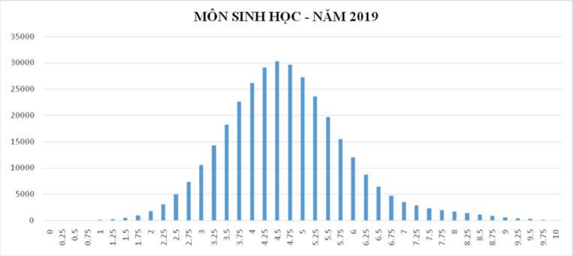
Z-table is used to find probability.

The SAT score follows the normal distribution with mean 1150 and standard deviation 150. You take a SAT with a score 1400, what is your SAT percentile?

Histogram

3. Sinh học

Điểm	**0**	**0.25**	**0.5**	**0.75**	**1**	**1.25**	**1.5**	**1.75**	**2**	**2.25**	**2.5**	**2.75**	**3**	**3.25**	**3.5**	**3.75**	**4**	**4.25**	**4.5**	**4.75**	**5**
Số lượng	0	2	9	17	70	191	455	946	1781	2972	4936	7311	10532	14233	18150	22546	26171	29056	30279	29624	27257
Điểm	**5.25**	**5.5**	**5.75**	**6**	**6.25**	**6.5**	**6.75**	**7**	**7.25**	**7.5**	**7.75**	**8**	**8.25**	**8.5**	**8.75**	**9**	**9.25**	**9.5**	**9.75**	**10**	
Số lượng	23613	19672	15461	11964	8641	6404	4704	3483	2850	2316	1893	1670	1333	1073	806	574	406	256	134	39	



Tổng số thí sinh	333830
Điểm trung bình	4.68
Điểm trung vị	4.50
Số thí sinh đạt <=1 điểm	98
Số thí sinh đạt điểm dưới trung bình (<5 điểm)	199281 (59.70%)
Điểm số có nhiều thí sinh đạt nhất	4.50

Central limit theorem

Regardless of the population distribution model, as the sample size increases, the sample mean tends to be normally distributed around the population mean, and its standard deviation shrinks as n increases.

Demo

Multivariate Gaussian distribution

For a D-dimensional vector x, the multivariate Gaussian distribution takes the form

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

μ is a D-dimensional mean vector, Σ is a D $\times$ D covariance matrix, and $|\Sigma|$ denotes the determinant of Σ

Multivariate Gaussian distribution (cont.)

Set

$$\triangle^2 = (x-\mu)^T\Sigma^{-1}(x-\mu)$$
$$= -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$$

It is a quadratic form of Gaussian distribution.

Multivariate Gaussian distribution (cont.)

Consider eigenvalues and eigenvectors of Σ

$$\Sigma u_i = \lambda_i u_i, \text{i} = 1,\ldots,\text{D}$$

Because Σ is a real, symmetric matrix, its eigenvalues will be real and its eigenvectors form an orthonormal set.

$$\Sigma = \sum_{i=1}^{D} \lambda_i u_i u_i^T \Rightarrow \Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i} u_i u_i^T$$

So that

$$\begin{aligned}\triangle^2 = (x-\mu)^T \Sigma^{-1} (x-\mu) &= \sum_{i=1}^{D} \frac{1}{\lambda_i} (x-\mu)^T u_i u_i^T (x-\mu) \\ &= \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}, \text{ with } y_i = u_i^T (x-\mu)\end{aligned}$$

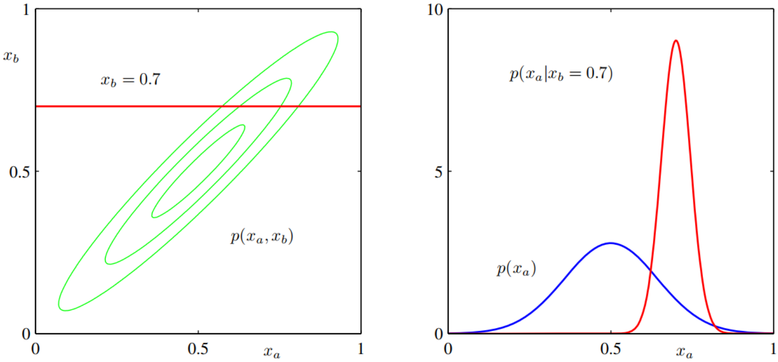
$$|\Sigma|^{1/2} = \prod_{j=1}^{D} \lambda_j^{1/2}$$

Multivariate Gaussian distribution (cont.)

$$
\begin{aligned}
p(y) &= \prod_{j=1}^{D} \frac{1}{(2\pi\lambda_j)}^{1/2} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\} \\
\Rightarrow \int_{-\infty}^{\infty} p(y)dy &= \prod_{j=1}^{D} \int_{-\infty}^{\infty} \frac{1}{(2\pi\lambda_j)}^{1/2} \exp\left\{-\frac{y_j^2}{2\lambda_j}\right\} dy_j \\
&= 1
\end{aligned}
$$

If two sets of variables are jointly Gaussian, then the conditional distribution of one set conditioned on the other is again Gaussian. Similarly, the marginal distribution of either set is also Gaussian.

Multivariate Gaussian distribution (cont.)



Hình 3: The plot on the left shows the contours of a Gaussian distribution $p(x_a, x_b)$ over two variables, and the plot on the right shows the marginal distribution $p(x_a)$ (blue curve) and the conditional distribution $p(x_a|x_b = 0.7)$ (red curve)

Conditional Gaussian Distribution

Suppose x is a D-dimensional vector with Gaussian distribution $\mathcal{N}(x|\mu, \Sigma)$ and that we partition x into two disjoint subsets x_a and x_b

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$$

We also define corresponding partitions of the mean vector μ given by

$$\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$$

and of the covariance matrix Σ given by

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \Rightarrow A = \Sigma^{-1} = \begin{pmatrix} A_{aa} & A_{ab} \\ A_{ba} & A_{bb} \end{pmatrix}$$

Σ is symmetric so Σ_{aa} and Σ_{bb} are symmetric while $\Sigma_{ab} = \Sigma_{ba}^T$

We are looking for conditional distribution $p(x_a|x_b)$

Conditional Gaussian Distribution (cont.)

We have

$$
\begin{aligned}
-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) &= -\frac{1}{2}(x-\mu)^T A(x-\mu) \\
&= -\frac{1}{2}(x_a-\mu_a)^T A_{aa}(x_a-\mu_a) - \frac{1}{2}(x_a-\mu_a)^T A_{ab}(x_b-\mu_b) \\
&\quad - \frac{1}{2}(x_b-\mu_b)^T A_{ba}(x_a-\mu_a) - \frac{1}{2}(x_b-\mu_b)^T A_{bb}(x_b-\mu_b) \\
&= -\frac{1}{2}x_a^T A_{aa}^{-1} x_a + x_a^T(A_{aa}\mu_a - A_{ab}(x_b-\mu_b)) + const
\end{aligned}
$$

It is quadratic form of x_a hence conditional distribution $\mathrm{p}(x_a|x_b)$ will be Gaussian, because this distribution is characterized by its mean and its variance. Compare with Gaussian disitrubtion

$$\triangle^2 = -\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu + const$$

Conditional Gaussian Distribution (cont.)

$$\Sigma_{a|b} = A_{aa}^{-1}$$
$$\mu_{a|b} = \Sigma_{a|b}(A_{aa}\mu_a - A_{ab}(x_b - \mu_b)) = \mu_a - A_{aa}^{-1}A_{ab}(x_b - \mu_b)$$

By using Schur complement,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CMD^{-1} & D^{-1}CMBD^{-1}, M = (A - BD^{-1}C)^{-1} \end{pmatrix}$$
$$\Rightarrow A_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$$
$$A_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$$

As a result

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$$
$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$
$$\Rightarrow p(x_a|x_b) = \mathcal{N}(x_{a|b}|\mu_{a|b}, \Sigma_{a|b})$$

Marginal Gaussian distribution

The margianl distribution given by

$$p(x_a) = \int p(x_a, x_b) dx_b$$

We need to integrate out x_b by looking the quadratic form related to x_b

$$-\frac{1}{2}x_b^T A_{bb} x_b + x_b^T m = -\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m) + \frac{1}{2}m^T A_{bb}^{-1} m$$

with $m = A_{bb}\mu_b - A_{ba}(x_a - \mu_a)$

We can integrate over unnormalized Gaussian

$$\int \exp\left\{-\frac{1}{2}(x_b - A_{bb}^{-1}m)^T A_{bb}(x_b - A_{bb}^{-1}m)\right\} dx_b$$

The remaining term

$$-\frac{1}{2}x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})x_a + x_a^T(A_{aa} - A_{ab}A_{bb}^{-1}A_{ba})^{-1}\mu_a + const$$

Similarly, we have

$$\mathbb{E}[x_a] = \mu_a$$
$$cov[x_a] = \Sigma_{aa}$$
$$\Rightarrow p(x_a) = \mathcal{N}(x_a|\mu_a, \Sigma_{aa})$$