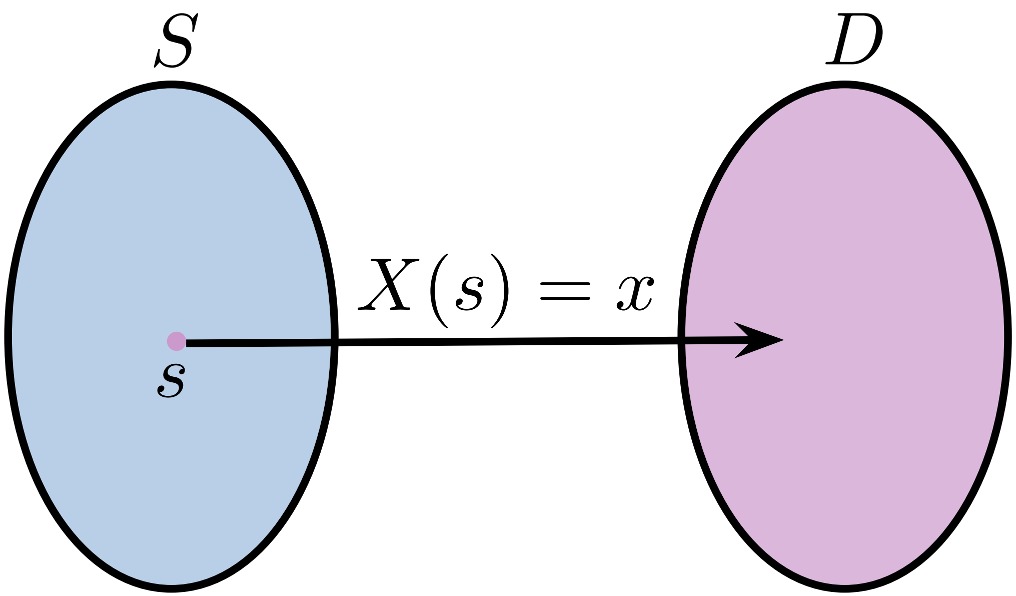
통계분석
Statistical Analysis

Discrete Random Variable

- The **number of heads** when an unfair coin is tossed n times
- The **number of visits** to some particular website

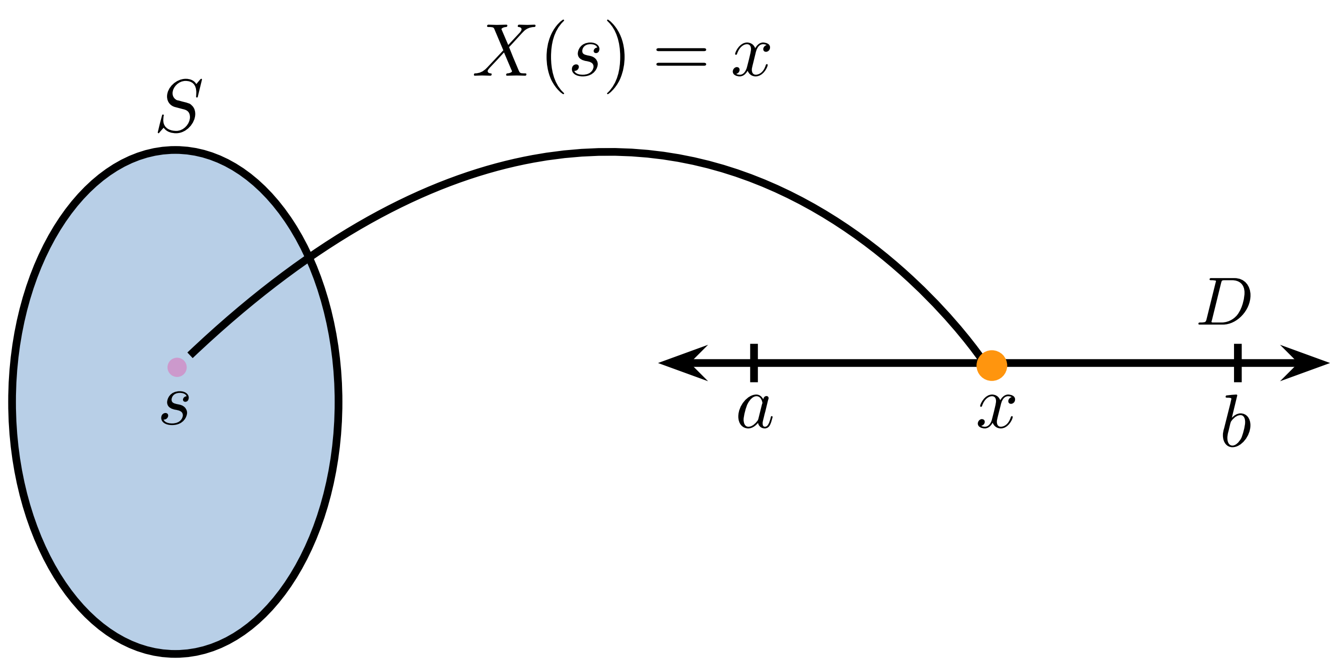


$$x \in D = \{ \text{ a countable set of numbers } \}$$

Continuous Random Variable

- The **length of time** that it takes to draw money from ATM
- The **weight** of a big mac burger sold at McDonald's in Time World

These numbers are ***not discrete*** ones such as integers.
They could be one number in some range of real number ***[a, b]***.

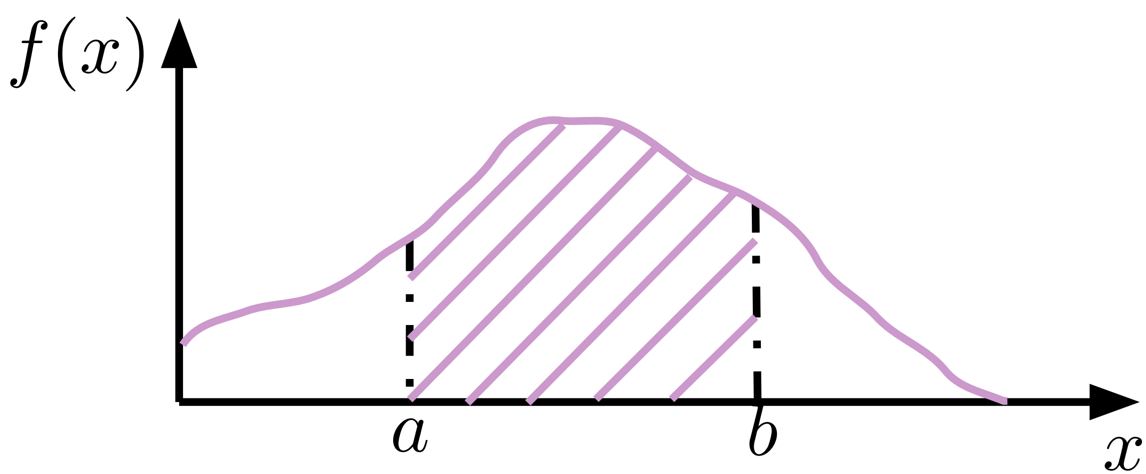


Probability Density Distribution

For a continuous random variable ***X*****, Probability distribution (probability density function)** is a non-negative function *f(x)* such that for any two numbers *a* and *b*

$$P(a \le X \le b) = \int_a^b f(x)dx$$

$P(a \le X \le b) =$ The probability that X has a value in the interval [a, b]



Probability Density Distribution

For a continuous random variable ***X*****, Probability distribution (probability density function)** is a non-negative function *f(x)* such that for any two numbers *a* and *b*

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

$P(a \leq X \leq b) =$ The probability that X has a value in the interval [a, a+dx]

dx = very small, infinitesimal

P(x) = f(x)dx when x is in the interval [a, a+dx]
f(x) = P(x)/dx (dimension, unit is not probability)

Probability density is not equal to probability.

Mathematical Expectation

- **Expected or mean value** of r.v. X with pdf f(x)

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

- **Expected or mean value of some function** h(X) of r. v. X with pdf f(x)

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx$$

- **Variance** of X with pdf f(x)

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx = E[(X-\mu)^2]$$
$$= E(X^2) - [E(X)]^2$$

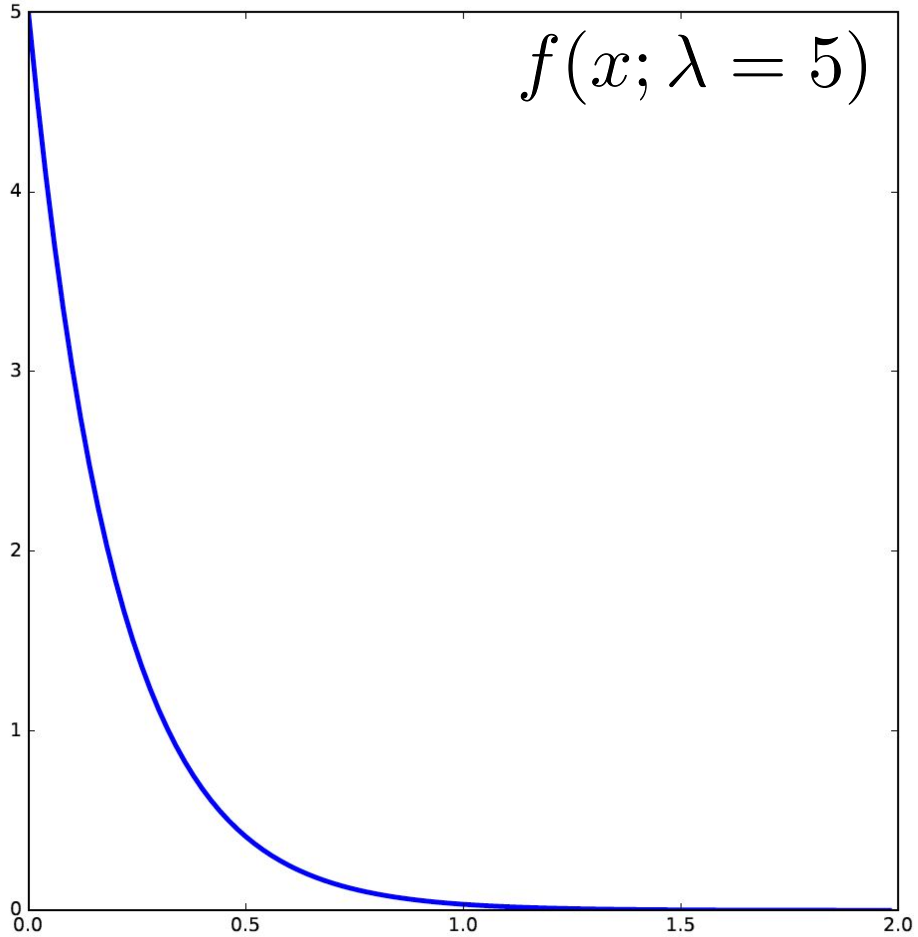
- **Standard deviation (SD)** of X

$$\sigma_X = \sqrt{V(X)}$$

Continuous Distributions

1. Exponential Distribution

$$f(x; \lambda) = \lambda \exp(-\lambda x) \quad x \geq 0$$



2. Gamma Distribution

For $\alpha > 0, \beta > 0$

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \geq 0$$

When $\beta = 1$,

$$f(x; \alpha, 1) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \quad x \geq 0$$

Standard Gamma distribution

$\Gamma(\alpha)$??

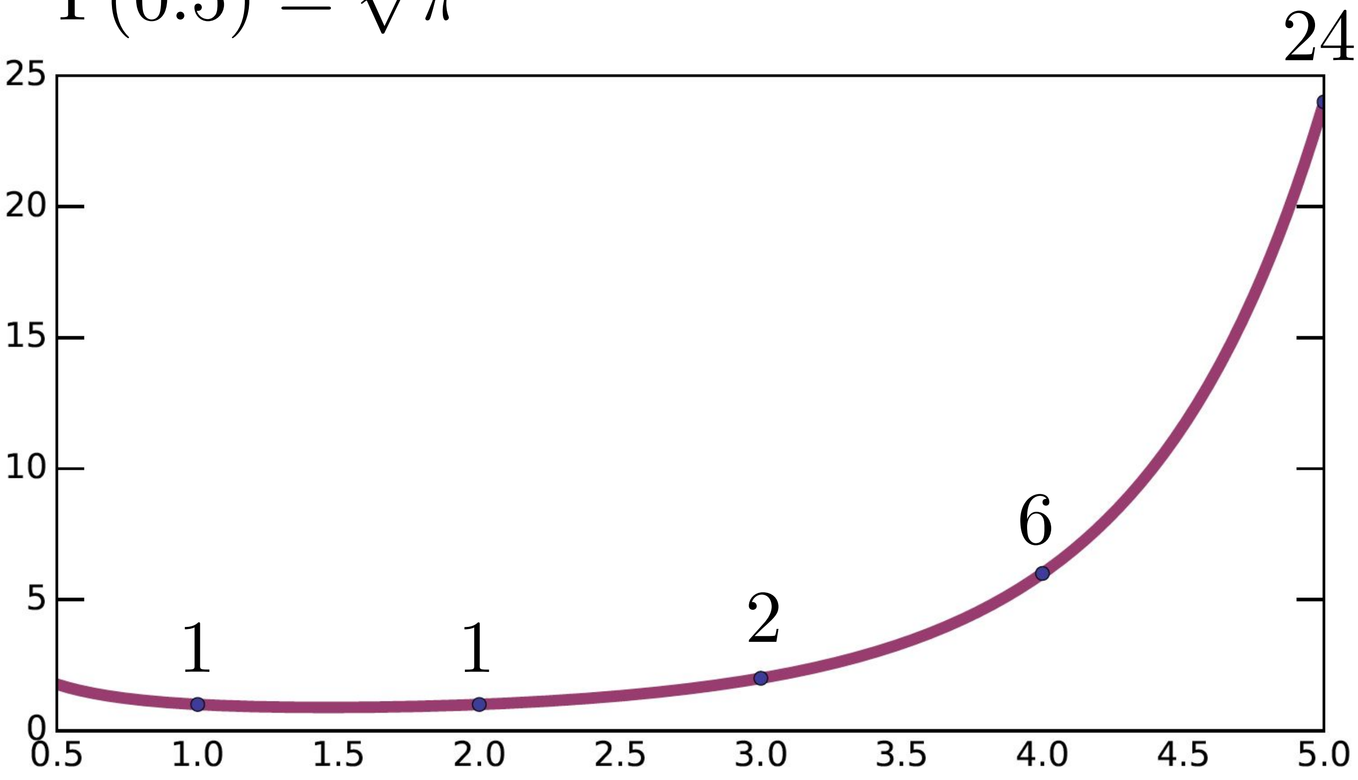
Gamma Function

For $\alpha > 0$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

For positive integer $n, \Gamma(n) = (n-1)!$

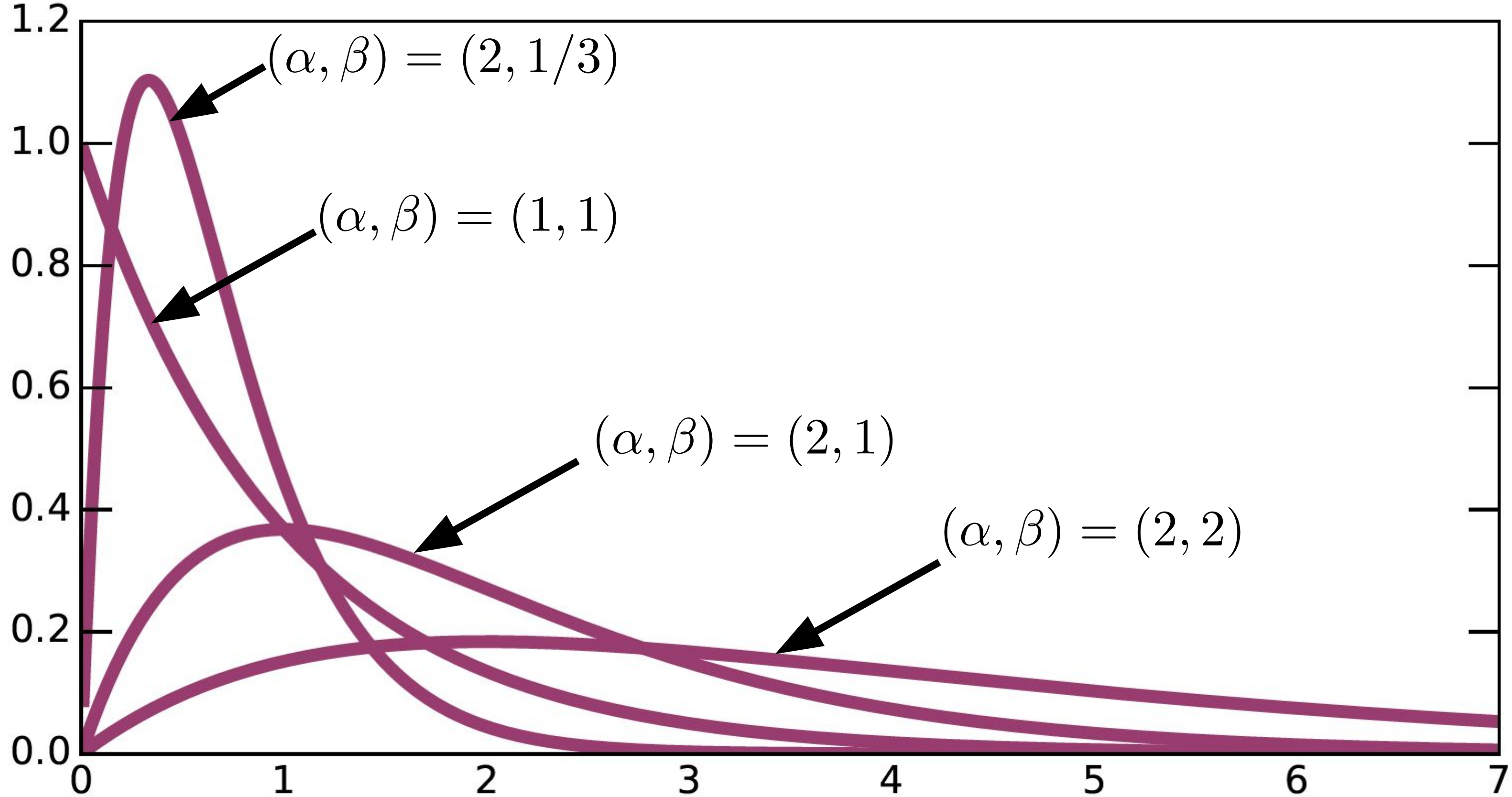
$\Gamma(0.5) = \sqrt{\pi}$



Gamma Distribution

For $\alpha > 0, \beta > 0$

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \geq 0$$



Gamma Distribution ~ Exponential Distribution

For $\alpha > 0, \beta > 0$

$$f(x; \alpha, \beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \quad x \geq 0$$

If $\alpha = 1, \beta = 1/\lambda$

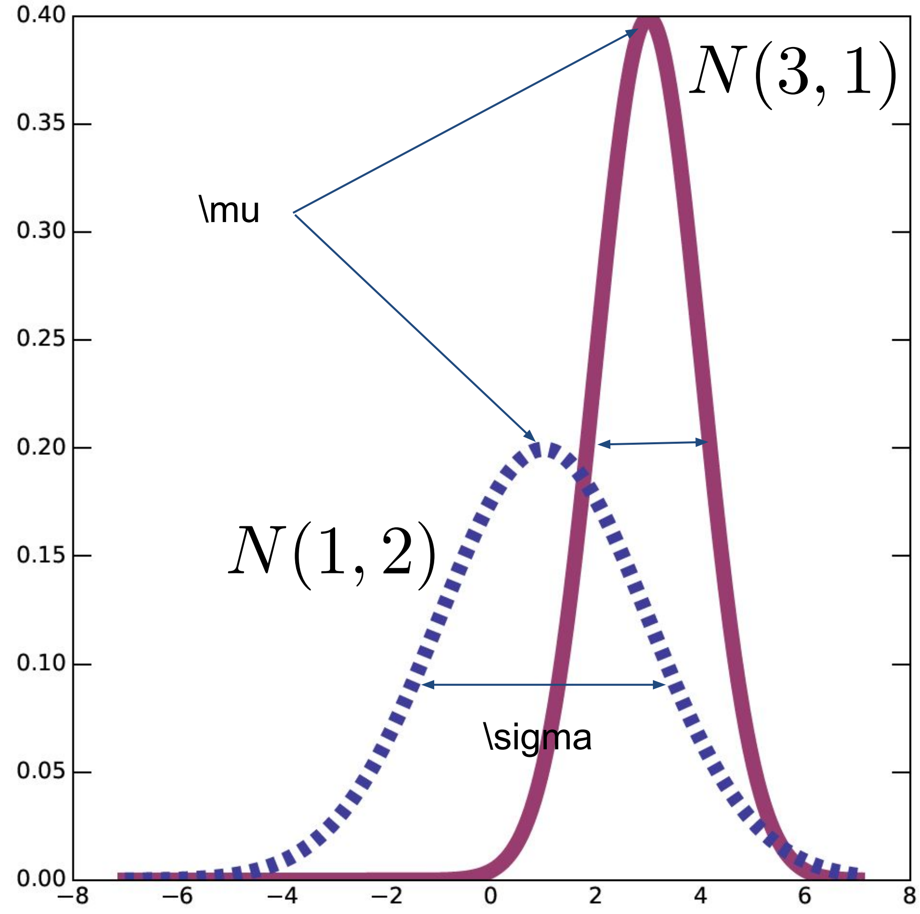
$$f(x; 1, 1/\lambda) = \lambda e^{-\lambda x} \quad x \geq 0$$

Exponential distribution

3. Normal Distribution (ND)****

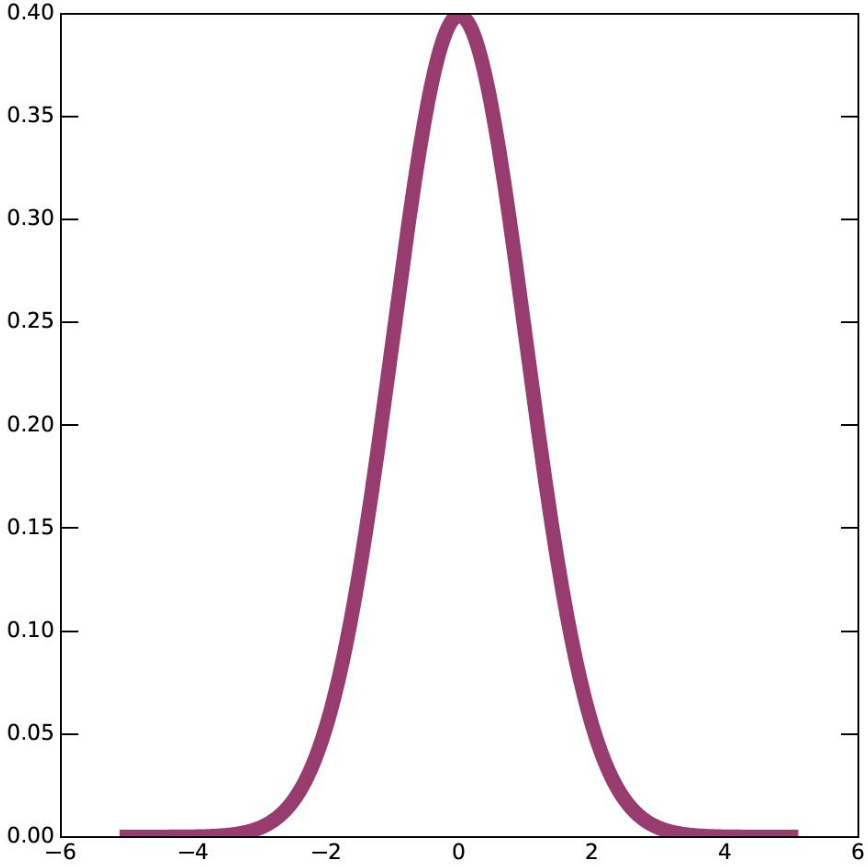
$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \equiv N(\mu,\sigma^2)$$

Gaussian Function



Standard Normal Distribution

$$N(\mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

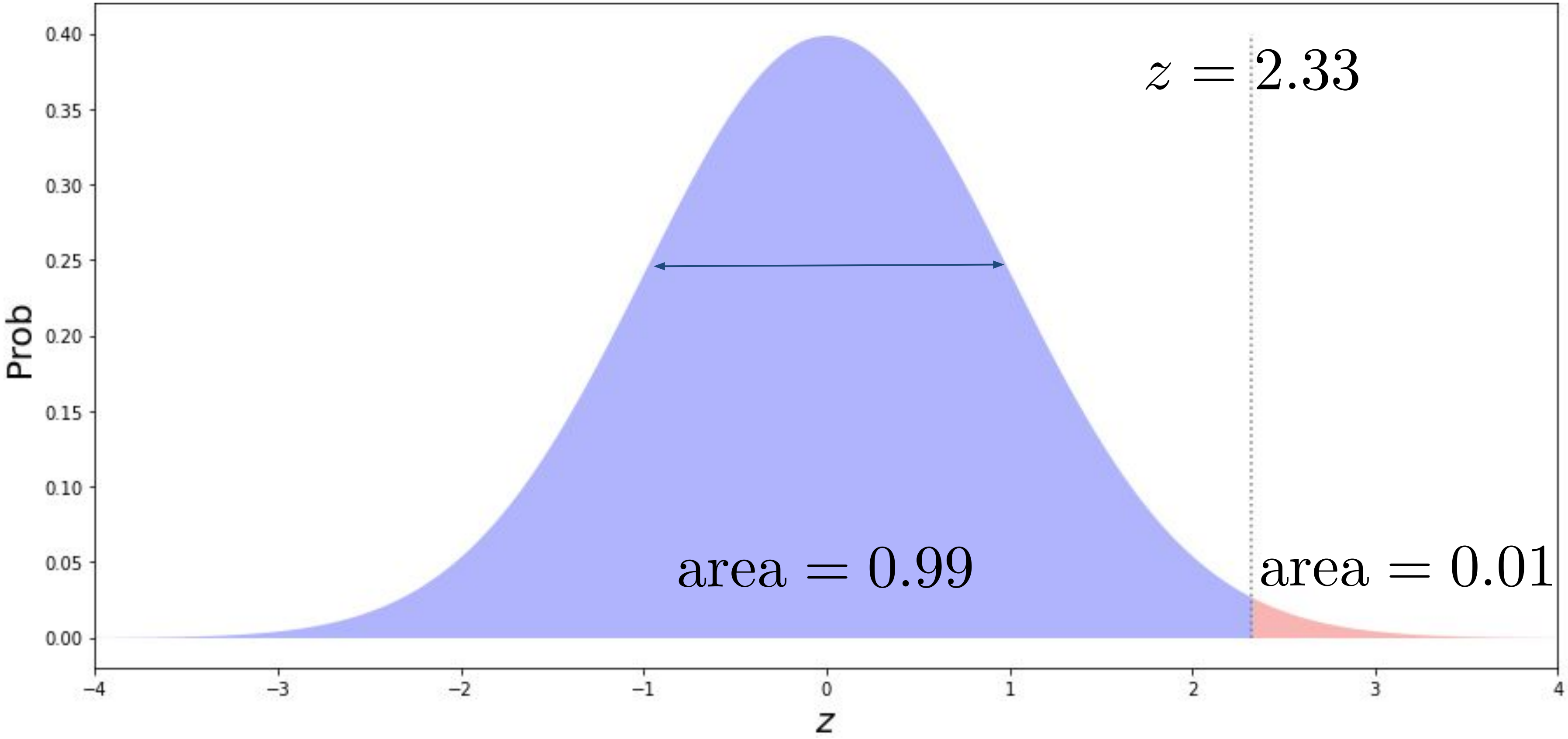


We call the random variable for the standard normal distribution as ***z***.

Percentiles of Standard ND

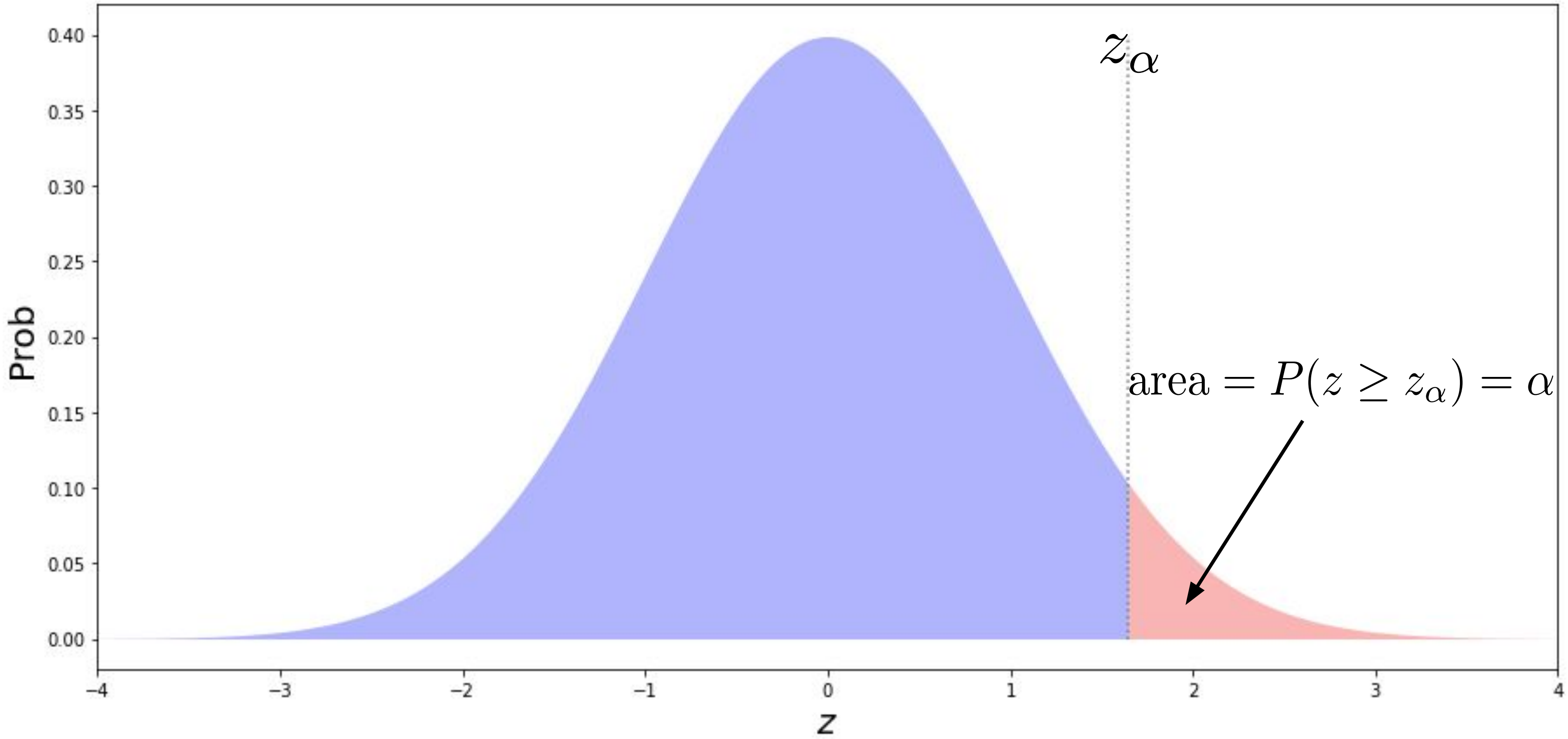
The 99th percentile of Standard Normal Distribution: $z = 2.33$

$$N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$



Notation for z critical values

For $\alpha (< 1.0)$ z_α: the z value such that $P(z \geq z_\alpha) = \alpha$

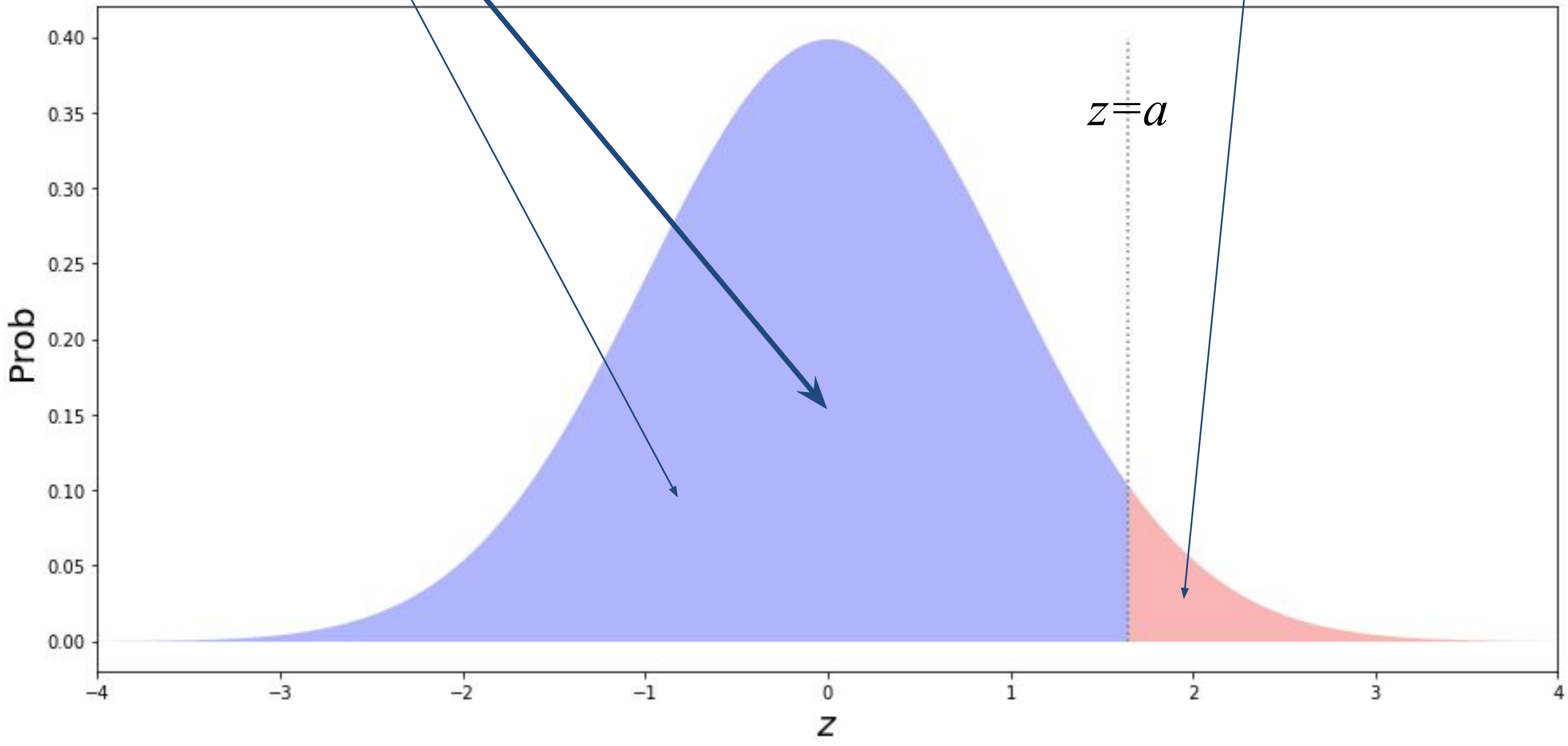


$z_\alpha = 100(1 - \alpha)$percentile of standard normal distribution

Percentiles and Critical values

For $N(\mu = 0, \sigma = 1)$

$$\Phi(a) \equiv \int_{-\infty}^{a} N(z; 0, 1) dz = P(z \leq a)$$



Standardization of Normal Distribution

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \longrightarrow N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

$$z = \frac{x-\mu}{\sigma} \qquad dz = \frac{1}{\sigma}dx$$

$$\int N(x;\mu,\sigma^2)dx = \int N(z;0,1)dz$$

- Any normal distribution can be described by using characteristics of the standard normal distribution under this transformation.

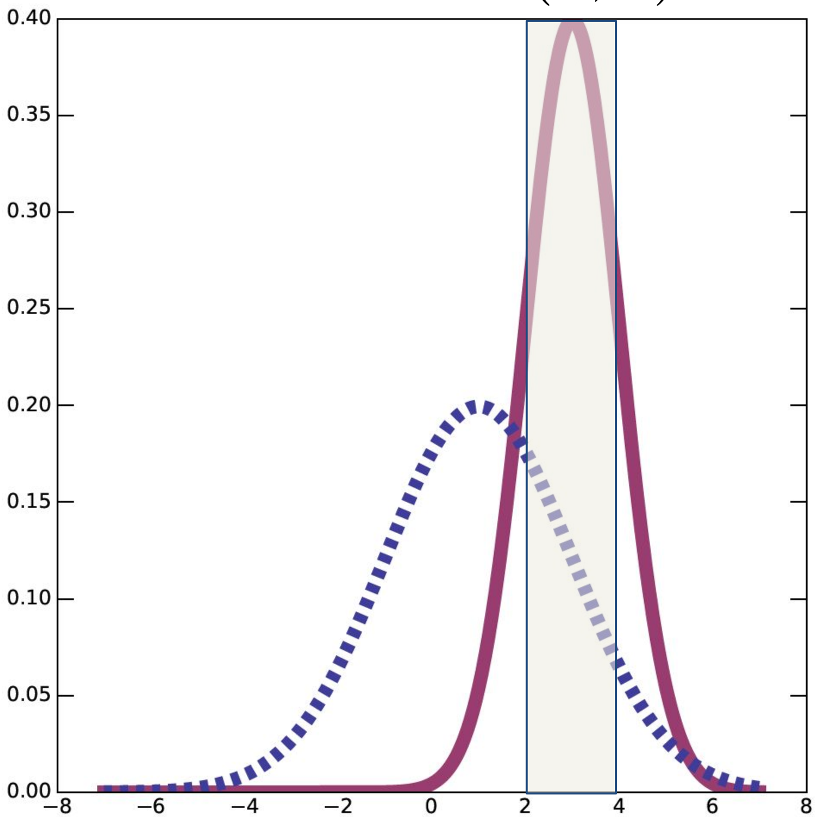
- Step 1. Transform from your original normal distribution to standard one.
- Step 2. Calculate statistical numbers by using standard normal distribution.
- Step 3. Transform back to the original normal distribution.

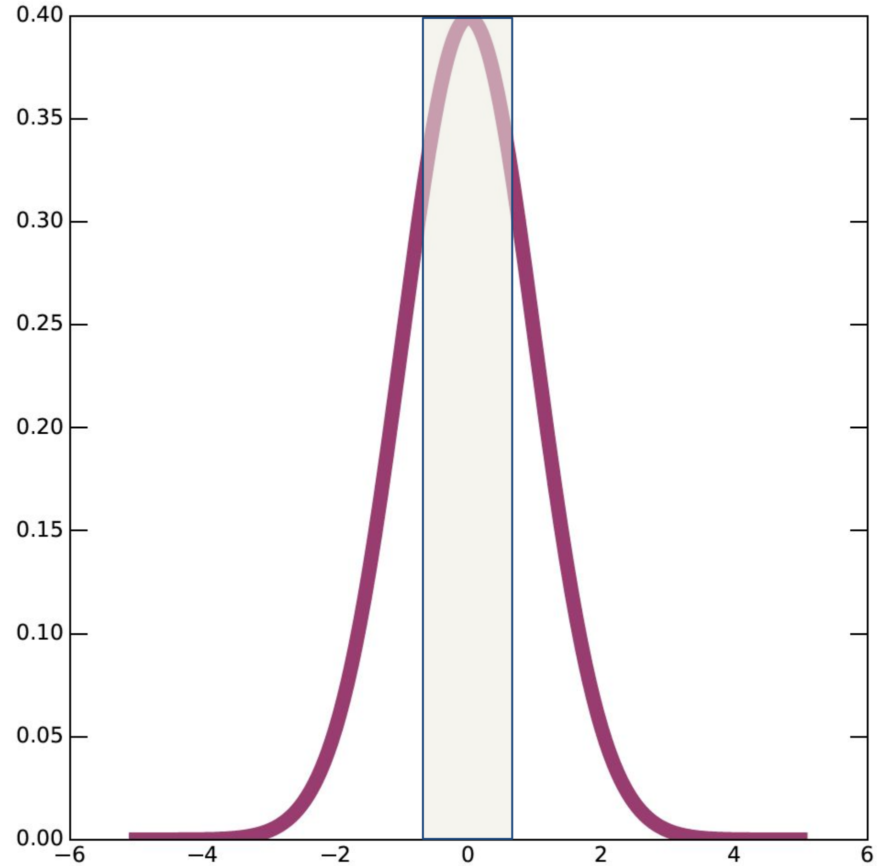
Standardization of Normal Distribution

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \rightarrow N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

$$z = \frac{x-\mu}{\sigma} \qquad dz = \frac{1}{\sigma}dx$$

$N(3,1)$





Standardization of Normal Distribution

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \longrightarrow N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

$$z = \frac{x-\mu}{\sigma} \qquad dz = \frac{1}{\sigma}dx$$

What is the area of N(3,1) when x is from 2 to 4?

$$(2,4) \text{ for } N(3,1) \xrightarrow{z = x - 3} (-1,1) \text{ for } N(0,1)$$

Read Area of N(0,1) in (-1,1) from some table.

$$S = \int_{-1}^{1} N(z; 0, 1) dz$$

$$\int_{2}^{4} N(x; \mu, \sigma^2) dx = \int_{-1}^{1} N(z; 0, 1) dz$$

Standardization of Normal Distribution

$$X \sim N(\mu, \sigma^2)$$

Random variable X has a normal distribution with mean μ and variance σ^2.

- **Standardization** $Z = (X - \mu)/\sigma \qquad Z \sim N(0, 1)$

$$P(a \le X \le b) = P\left(\frac{(a-\mu)}{\sigma} \le Z \le \frac{(b-\mu)}{\sigma}\right) \quad Z \sim N(0,$$

$$X \sim N(\mu, \sigma^2)$$

$$= \Phi\left(\frac{(b-\mu)}{\sigma}\right) - \Phi\left(\frac{(a-\mu)}{\sigma}\right)$$

$$P(X \le a) = \Phi\left(\frac{(a-\mu)}{\sigma}\right)$$

Percentiles of Arbitrary Normal Distribution

Standardization

$$X \sim N(\mu, \sigma^2) \xrightarrow[Z = (X-\mu)/\sigma]{} Z \sim N(0,1)$$

$$P(X \leq a) = \Phi\left(\frac{(a-\mu)}{\sigma}\right)$$

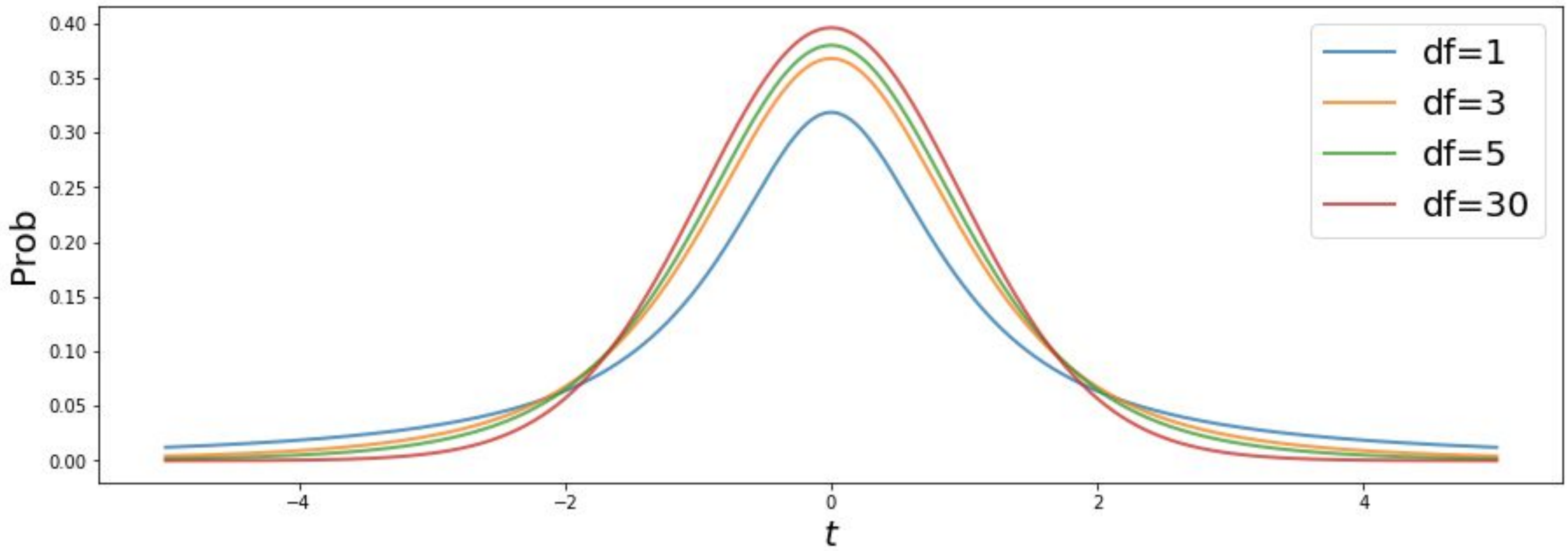
$$(100p)\text{th percentile for } N(\mu, \sigma^2) = \mu + [(100p)\text{th percentile for } N(0,1)]\,\sigma$$

Student's *t* distribution

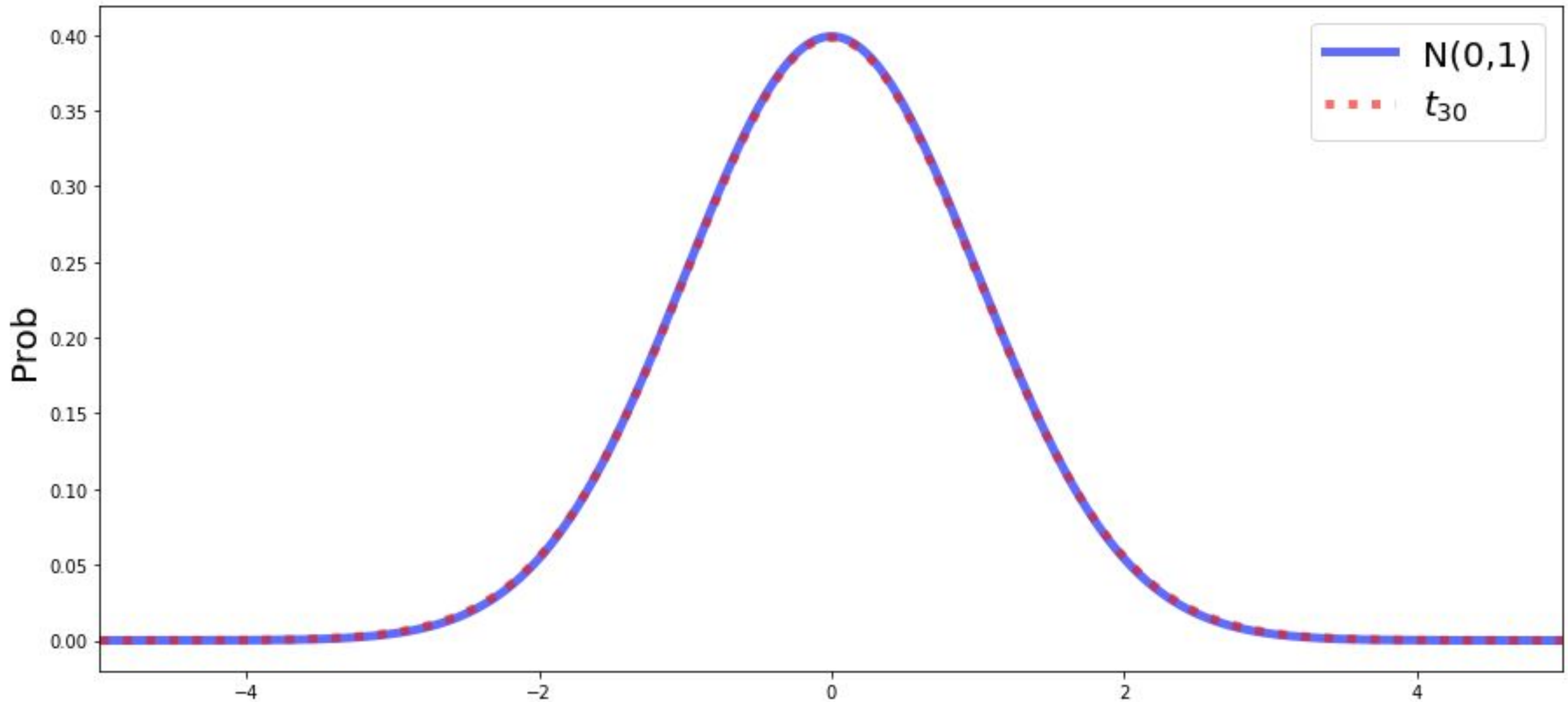
For random variable ***T=t***

$$f(t;\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} = t_\nu$$

$$\nu = \text{degree of freedom}$$



Student's *t* Distribution vs Standard Normal Distribution



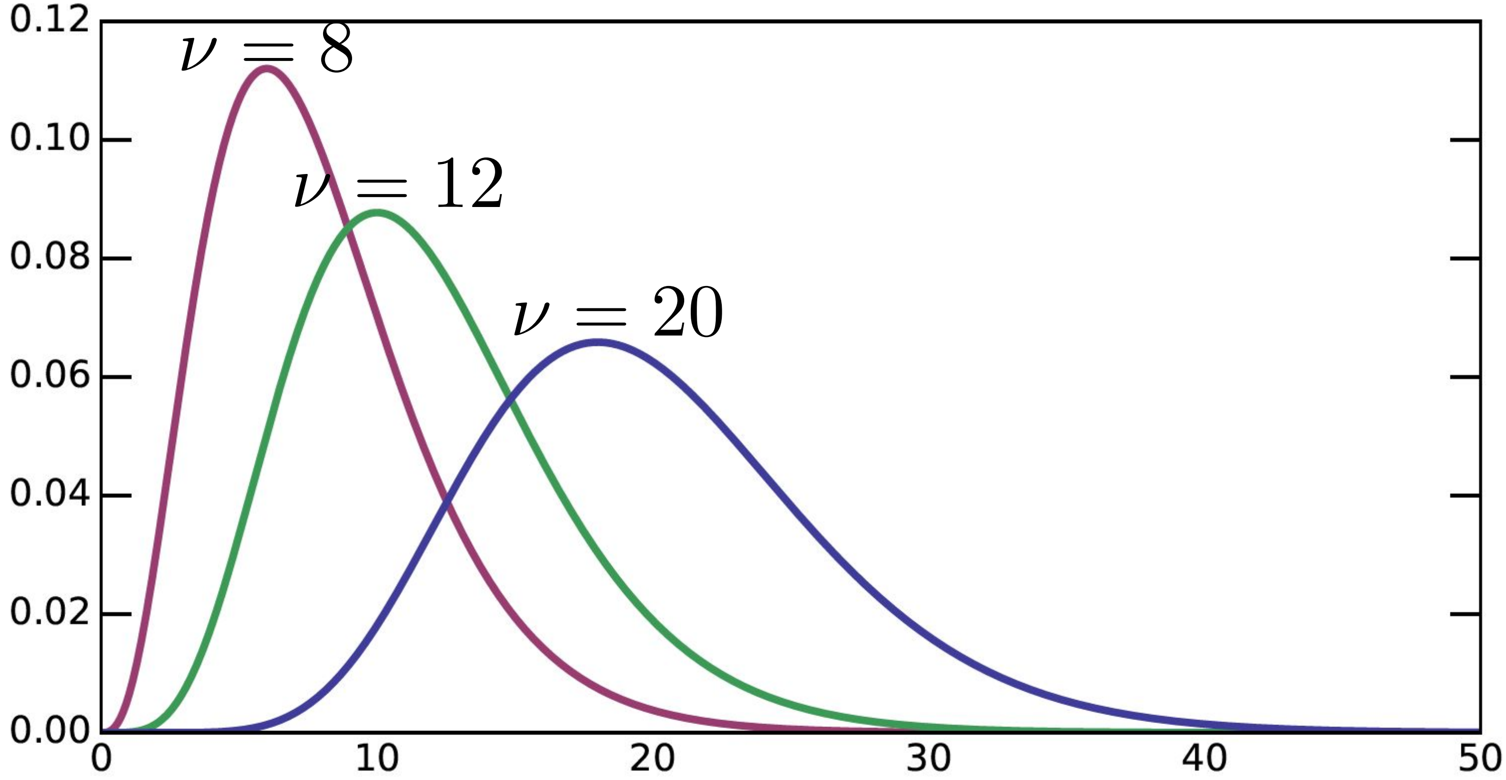
- If degree of freedom increases up to some value, t distribution approaches the standard normal distribution.
- (Actual) We need to use t distribution. Due to similarity between t distribution and standard normal distribution, we might approximately use S.N.D.
- We will see this approximation in the inferential statistics.

Chi-squared Distribution

For random variable ***X*=*x***

$$f(x;\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} \quad x \geq 0$$

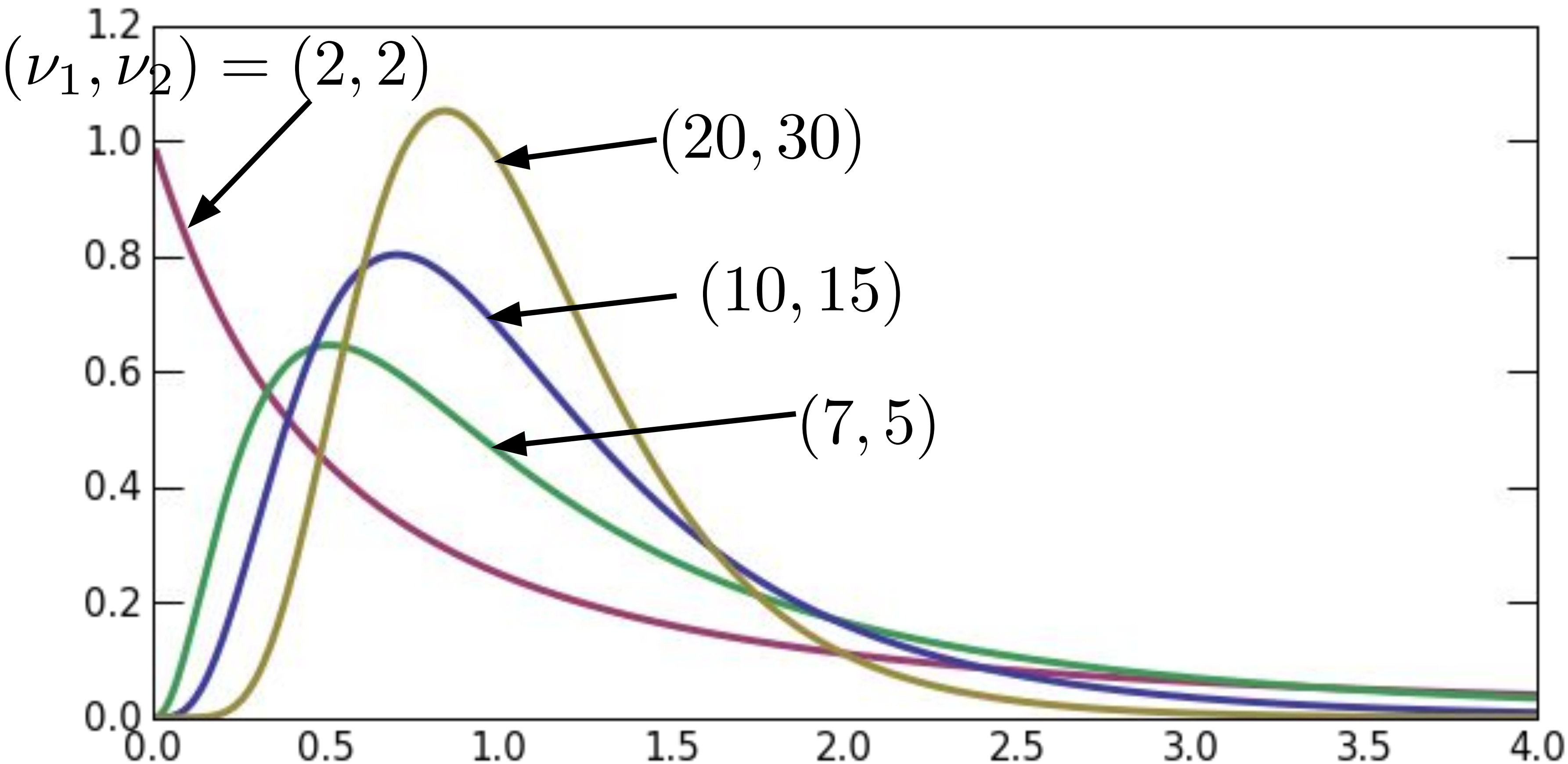
$\nu =$ degree of freedom



F-distribution

For r.v. $X = x$ with two degrees of freedom ν_1, ν_2

$$f(x;\nu_1,\nu_2) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\nu_1/2)\Gamma(\nu_1/2)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{x^{(\nu_1-2)/2}}{\left(1+\frac{\nu_1}{\nu_2}x\right)^{(\nu_1+\nu_2)/2}}$$



Continuous Distribution

- Normal distribution / Standard normal distribution
- Student's *t* distribution
- Chi-squared distribution
- F-distribution

These distributions will be used when we study

1. **sampling distribution**
2. **hypothesis testing**.
3. **Estimation**