# COMPSCI 240: Reasoning Under Uncertainty 

Nic Herndon and Andrew Lan<br>University of Massachusetts at Amherst

Spring 2019

Lecture 15: Normal/Gaussian Random Variables

## Normal (Gaussian) Random Variables

- A normal random variable $X$ is a continuous random variable with probability density function:

$$
f_{X}(x)=\frac{1}{\sqrt{2 \pi \sigma^{2}}} \cdot e^{-\frac{1}{2 \sigma^{2}}(x-\mu)^{2}}
$$

- The range of this random variable is $\mathcal{X}=(-\infty, \infty)$.
- The parameter $\sigma$ must be strictly greater than 0 .
- The parameter $\mu$ can be any real value.
- Often also abbreviated as $X \sim \mathcal{N}\left(\mu, \sigma^{2}\right)$.
- Normal random variables have extremely important theoretical properties and are a default choice for modeling problems involving continuous measurements.


## Why Study Normal RV?

- Normal RV plays an important role in a broad disciplines, including but not limited to computer science, engineering, physical, and statistical context.
- Few examples related to computer science include linear regression and Gaussian Process.
- The normal random variable is a convenient tool to approximate various types of phenomena (or observations), which allows us to derive mathematically tractable solutions.
- The key fact is that the sum of a large number of independent and identically distributed (not necessarily normal) random variables has an approximately normal behavior (Central Limit Theorem).

Normal (Gaussian) Random Variables


## Mean, Variance, and CDF

- The mean and variance can be calculated to be

$$
E[X]=\mu \text { and } \operatorname{var}(X)=\sigma^{2}
$$

- Its CDF is defined as

$$
F_{X}(x)=P(X \leq x)=\frac{1}{\sigma \sqrt{2 \pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d x
$$

- The probability mass of an interval $[a, b]$ is the definite integral:

$$
\begin{aligned}
P(a<X<b) & =\frac{1}{\sigma \sqrt{2 \pi}} \int_{a}^{b} e^{-\frac{(x-\mu)^{2}}{2 \sigma^{2}}} d x \\
& =F_{X}(b)-F_{X}(a)
\end{aligned}
$$

## Normal (Gaussian) Random Variables



## Preserving the Normality (or Gaussianity)

- Let $X$ be a normal random variable with mean $\mu$ and variance $\sigma^{2}$, and if $a \neq 0$ and $b$ are scalars, then a random variable

$$
Y=a X+b
$$

is also a normal random variable with

$$
E[Y]=a \mu+b \text { and } \operatorname{var}(Y)=a^{2} \sigma^{2} .
$$

## The Standard Normal Random Variable

- A normal random variable $X$ with $E[X]=0$ and $\operatorname{var}(X)=1$ is said to be a standard normal random variable.
- Its PDF can be simplified as

$$
f_{X}(x)=\frac{1}{\sqrt{2 \pi}} \cdot e^{-\frac{1}{2} x^{2}}
$$



## The Standard Normal Random Variable

- Its CDF can be defined as

$$
\Phi(x)=P(X \leq x)=P(X<x)=\frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{x} e^{-\frac{1}{2} t^{2}} d t
$$

## The Standard Normal Table

| $\Phi(x)$ | 0 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0 | .5 | .50399 | .50798 | .51197 | .51595 | .51994 | .52392 | .5279 | .53188 | .53586 |
| .1 | .53983 | .5438 | .54776 | .55172 | .55567 | .55962 | .56356 | .56749 | .57142 | .575355 |
| .2 | .57926 | .58317 | .58706 | .59095 | .59483 | .59871 | .60257 | .60642 | .61026 | .61409 |
| .3 | .61791 | .62172 | .62552 | .6293 | .63307 | .63683 | .64058 | .64431 | .64803 | .65173 |
| .4 | .65542 | .6591 | .66276 | .6664 | .67003 | .67364 | .67724 | .68082 | .68439 | .68793 |
| .5 | .69146 | .69497 | .69847 | .70194 | .7054 | .70884 | .71226 | .71566 | .71904 | .7224 |
| .6 | .72575 | .72907 | .73237 | .73565 | .73891 | .74215 | .74537 | .74857 | .75175 | .7549 |
| .7 | .75804 | .76115 | .76424 | .7673 | .77035 | .77337 | .77637 | .77935 | .7823 | .78524 |
| .8 | .78814 | .79103 | .79389 | .79673 | .79955 | .80234 | .80511 | .80785 | .81057 | .81327 |
| .9 | .81594 | .81859 | .8121 | .82381 | .82639 | .82894 | .83147 | .83398 | .83646 | .83891 |
| 1 | .84134 | .84375 | .84614 | .84849 | .85083 | .85314 | .85543 | .85769 | .85993 | .86214 |
| 1.1 | .86433 | .8665 | .86864 | .87076 | .87286 | .87493 | .87698 | .879 | .881 | .88298 |
| 1.2 | .88493 | .88686 | .88877 | .89065 | .89251 | .89435 | .89617 | .89796 | .89973 | .90147 |
| 1.3 | .9032 | .9049 | .90658 | .90824 | .90988 | .91149 | .91309 | .91466 | .91621 | .91774 |
| 1.4 | .91924 | .92073 | .9222 | .92364 | .92507 | .92647 | .92785 | .92922 | .93056 | .93189 |
| 1.5 | .93319 | .93448 | .93574 | .93699 | .93822 | .93943 | .94062 | .94179 | .94295 | .94408 |
| 1.6 | .9452 | .9463 | .94738 | .94845 | .9495 | .95053 | .95154 | .95254 | .95352 | .95449 |
| 1.7 | .95543 | .95637 | .95728 | .95818 | .95907 | .95994 | .9608 | .96164 | .96246 | .96327 |
| 1.8 | .96407 | .96485 | .96562 | .96638 | .96712 | .96784 | .96856 | .96926 | .96995 | .97062 |
| 1.9 | .97128 | .97193 | .97257 | .9732 | .97381 | .97441 | .975 | .97558 | .97615 | .9767 |
| 2 | .97725 | .97778 | .97831 | .97882 | .97932 | .97982 | .9803 | .98077 | .98124 | .98169 |
| 2.1 | .98214 | .98257 | .983 | .98341 | .98382 | .98422 | .98461 | .985 | .98537 | .98574 |
| 2.2 | .9861 | .98645 | .98679 | .98713 | .98745 | .98778 | .98809 | .9884 | .9887 | .98899 |
| 2.3 | .98928 | .98956 | .98983 | .9901 | .99036 | .99061 | .99086 | .99111 | .99134 | .99158 |
| 2.4 | .9918 | .99202 | .99224 | .99245 | .99266 | .99286 | .99305 | .99324 | .99343 | .99361 |
| 2.5 | .99379 | .99396 | .99413 | .9943 | .99446 | .99461 | .99477 | .99492 | .99506 | .9952 |
| 2.6 | .99534 | .99547 | .9956 | .99573 | .99585 | .99598 | .99609 | .99621 | .99632 | .99643 |
| 2.7 | .99653 | .99664 | .99674 | .99683 | .99693 | .99702 | .99711 | .9972 | .99728 | .99736 |
| 2.8 | .99744 | .99752 | .9976 | .99767 | .99774 | .99781 | .99788 | .99795 | .99801 | .99807 |
| 2.9 | .99813 | .99819 | .99825 | .99831 | .99836 | .99841 | .99846 | .99851 | .99856 | .99861 |

## Standardizing a Normal Variable

- For a given normal random variable $X$ with mean $\mu$ and variance $\sigma^{2}$, you can standardize it by defining a new random variable $Y$ given by

$$
Y=\frac{X-\mu}{\sigma}
$$

- Since $Y$ is a form of $a X+b$, where $a=\frac{1}{\sigma}$ and $b=-\frac{\mu}{\sigma}$, we know that the normality of $Y$ is preserved.
- The mean and variance of $Y$ can be computed as

$$
\begin{gathered}
E[Y]=\frac{E[X]-\mu}{\sigma}=0 \text { and } \\
\operatorname{var}(Y)=\frac{\operatorname{var}(X)}{\sigma^{2}}=1
\end{gathered}
$$

## Example

- Question: The midterm for our CS240 class can be modeled as a normal random variable with a mean of $\mu=75 \%$ and standard deviation of $\sigma=10 \%$. What is the probability that a randomly chosen student has less than or equal to $80 \%$ of score?
- Answer: Let $X$ be the score. Then, we first normalize $X$ to have zero mean and unit variance.

$$
Y=\frac{X-75}{10}
$$

- Then,

$$
\begin{aligned}
P(X<80) & =P\left(Y<\frac{80-75}{10}\right)=P\left(Y<\frac{1}{2}\right) \\
& =\Phi\left(\frac{1}{2}\right) \\
& =0.69146
\end{aligned}
$$

## Example

- Question: What is the probability that a randomly chosen student has greater than or equal to $80 \%$ of score?
- Answer:

$$
P(X \geq 80)=1-P(X<80)=1-\Phi\left(\frac{1}{2}\right)=0.30854
$$

