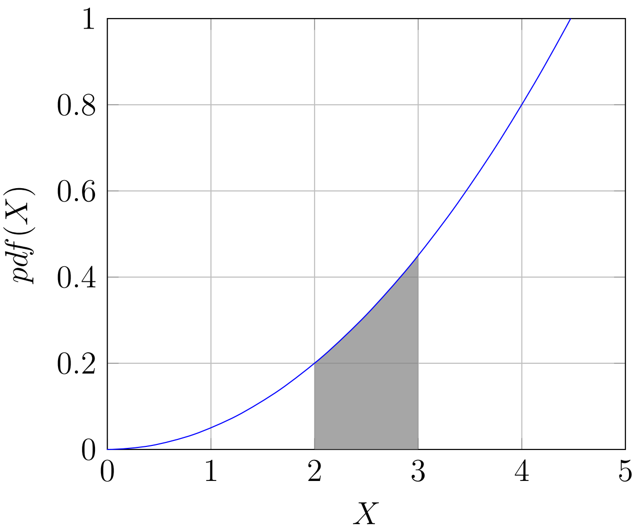
The Normal Distribution Unit

Andy Yan

January 2023

1 Introduction

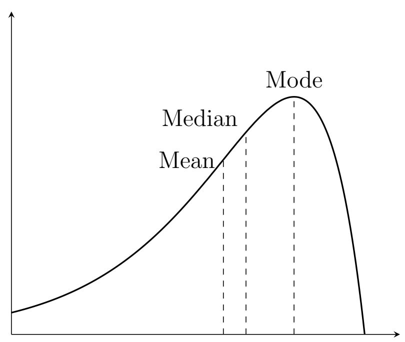
PDF (Probability Density Function)
The probability per unit change of the variable at a particular point.
Instead of the dependent variable being probability $P(X)$ it's replaced with $pdf(X)$.



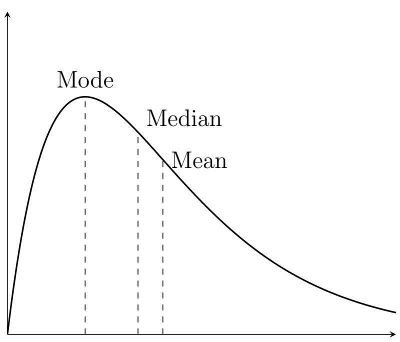
The shaded area represents $P(2 < X < 3)$. NOTE: $\leq$ and $<$ are the same

Terms
Unitary Condition: Total Probability $= 1$
Continuous Random Variable: No breaks in continuity (All values exist within its range)
Unimodal: One mode/median/mean on the curve
Bimodal: Two modes/medians/means on the curve

Negatively Skewed / Skewed Left

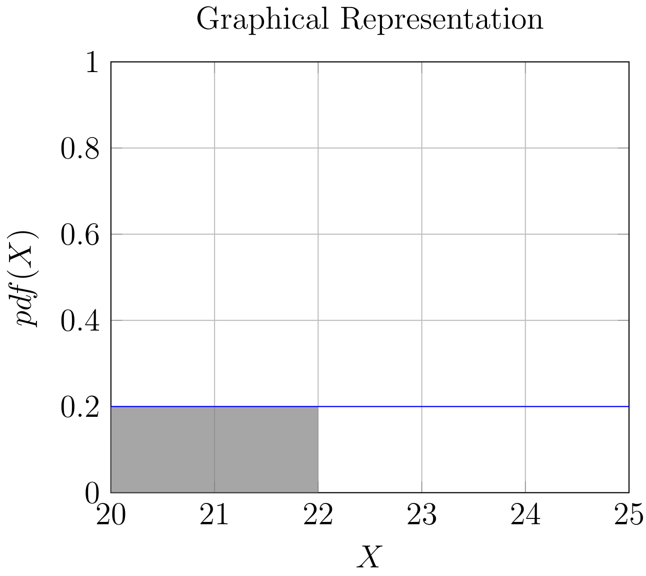


Positively Skewed / Skew Right



Uniform Distribution

Probability Density remains constant.



In this example, we are given $a = \frac{1}{5}$.

$$P(20 < X < 22) = 2 \cdot a = 2 \cdot \frac{1}{5} = \frac{2}{5}$$

Probability Density

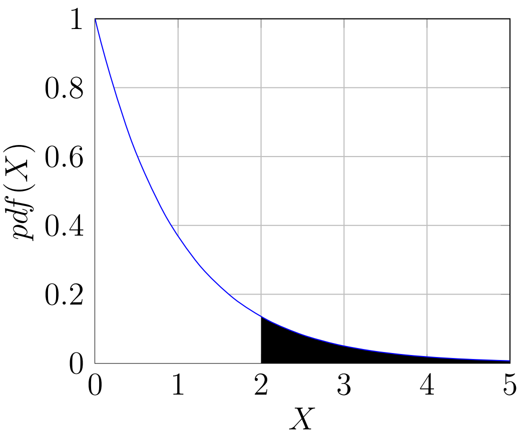
$$P(X < a) = ke^{-ka}$$

μ : average time
$k = \frac{1}{\mu}$: # of events per unit of time, ex. #per hr/min/sec

Exponential Distribution

Used to describe or find the probability related to time taken between 2 consecutive events that occurred.

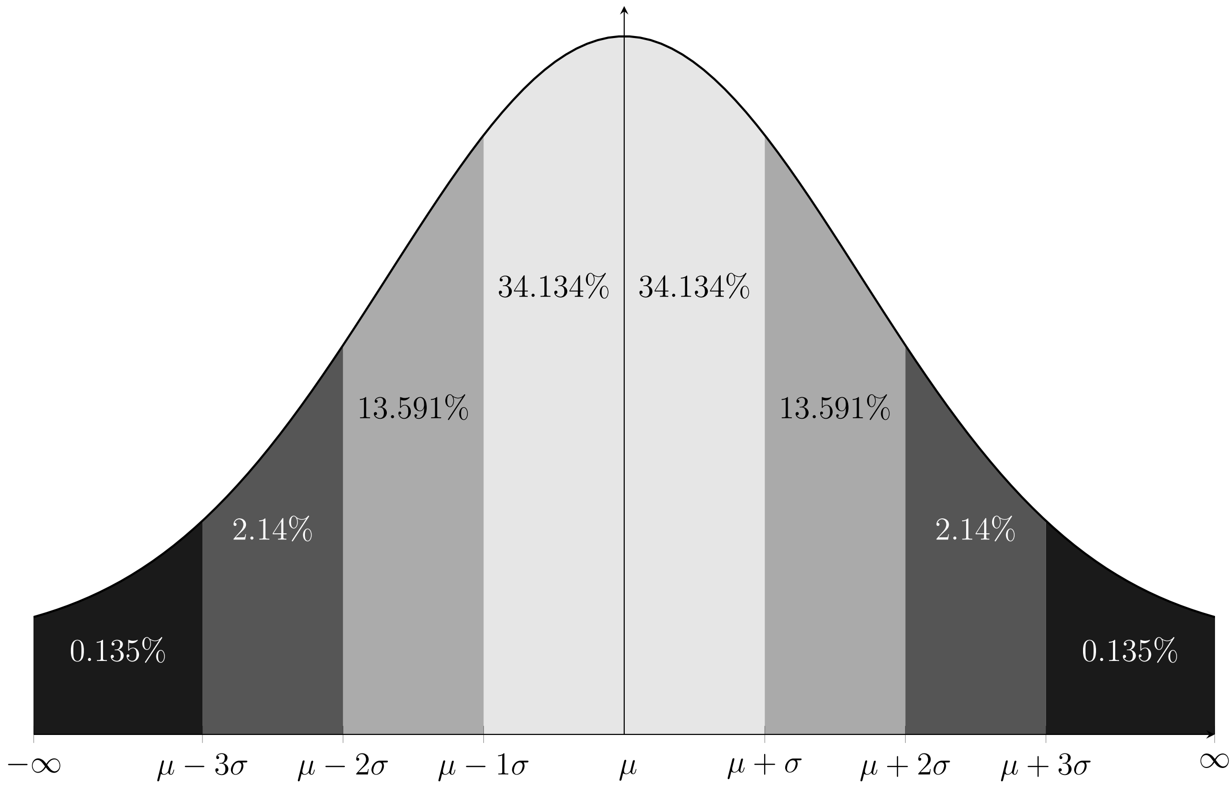
$$P(X < x) = 1 - e^{-kx} \text{ WHITE}$$
$$P(X > x) = e^{-kx} \text{ BLACK}$$



Example: if $x = 8, \mu = 10mins$

$$P(X < 8) = 1 - e^{-\frac{1}{10} \cdot 8} \approx 0.55mins$$

2 Normal Distribution



Properties of Normal Distribution

General Properties: Bell Shape, Symmetric, origin represents μ, total area under curve = 1
Distribution Type: Continuous Probability Distribution
y-axis represents pdf & area under curve represents probability

Formula:

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

x : random variable
μ : mean
σ : Standard Deviation

Exponent $(\frac{x-\mu}{\sigma})^2$ Properties:
When $x = \mu$, $f(x)$ has the maximum value. The greater $|x - \mu|$ is, the less $f(x)$ is.

Probability for Standard Normal Distribution

Each area interval of $\mu \pm Z\sigma$ $(Z = \{0, \pm 1, \pm 2, \pm 3\})$ represents a constant percentage which is the probability.

$$P(a \leq x \leq b) = \textbf{Sum of area intervals a to b} = \int_a^b \frac{e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2}}{\sigma\sqrt{2\pi}} dx \textit{ only to check answer}$$

Using Z-score $Z = \frac{x-\mu}{\sigma}$
If Z-score isn't whole: $x = \mu + Z(\sigma)$
Using Z-score table to find Z. Using positive table for positive Z-score and vice versa. Row represents hundredth digits, add the row and the column to find desired Z-score.

$$P(X < x) = P(Z < \frac{x-\mu}{\sigma})$$

$$P(Z < -x) = 1 - P(Z < x)$$

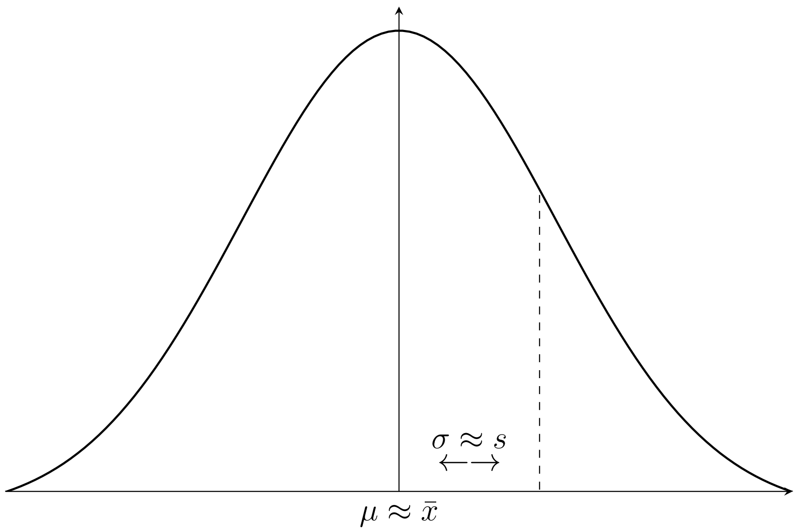
3 Normal Sampling and Modelling

Normal Sampling

1. The distribution of frequencies in the sample data tends to follow the bell-shaped curve as the underlying distribution.
2. The sample mean ($\bar{x}$) estimates the population mean (μ)
3. The sample standard deviation (s) estimates the population standard deviation (σ)
4. The larger the sample of a normal population, the more reliable the data will be in reflecting the underlying population

$$\bar{x} = \frac{\sum x}{n}$$

$$s = \sqrt{\frac{\sum x^2 - n(\bar{x})^2}{n-1}}$$



Normal Modelling

Being sensible: $\mu \pm Z\sigma$ shouldn't create any unreasonable data such as negative time, beyond capacity, etc.
Discrete data can sometimes be modelled by a normal distribution. The standard normal curve can be used to approximate the area under the curve which can be used to solve problems involving probability.

Continuity Correction

A continuity correction factor is used when you use a continuous probability distribution to approximate a discrete probability distribution.

- $P(X \leq a) \approx P(X < a + 0.5)$
- $P(X < a) \approx P(X < a - 0.5)$
- $P(X \geq a) \approx P(X > a - 0.5)$
- $P(X > a) \approx P(X > a + 0.5)$
- $P(X = a) \approx P(a - 0.5 < X < a + 0.5)$

Example:

$$P(X = 25) \rightarrow P(24.5 < X < 25.5)$$

4 Normal Approximation

REMEMBER the binomial distribution is perfectly symmetric if $p = \frac{1}{2}$, and has some skewness when not. The normal approximation works bets when p is close to $\frac{1}{2}$, and when n is large.

$$P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k}$$

The normal approximation is reasonable if both $np \geq 5$ and $n(1-p) \geq 5$.

Fora binomial random variable X:

$$\mu = np$$
$$\sigma^2 = np(1-p) = npq$$

Continuity Correction

Desired Probability	Normal Approximation
$P(X \geq x)$	$P(Z > \frac{(x-0.5)-\mu}{\sigma})$
$P(X > x)$	$P(Z > \frac{(x+0.5)-\mu}{\sigma})$
$P(X \leq x)$	$P(Z < \frac{(x+0.5)-\mu}{\sigma})$
$P(X < x)$	$P(Z < \frac{(x-0.5)-\mu}{\sigma})$

5 Repeated Sampling and Hypothesis Testing

Repeated Sampling

When repeated samples of the same size are drawn from a normal population, the sample means will be normally distributed with a mean equal to the population mean ($\mu_{\bar{x}} = \mu$). The distribution of sample means will be where n is the sample size:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

To find a Z-score for a sample mean:

$$P(\bar{x} < x) = P(Z < \frac{(x-0.5) - \mu_{\bar{X}}}{\sigma_{\bar{x}}})$$

Hypothesis Testing

A hypothesis test consists of a null hypothesis H_0 and an alternative hypothesis H_1. The null hypothesis is being challenged is for example H_0: $\mu = 45$. While we suspect H_1 :$\mu < 45$. To obtain a sample mean $\mu_{\bar{x}} = 44.5$ is very rare, and getting such will prove $\mu \neq 45$. Our next step is to establish a decision rule. The significance level α is the probability threshold for us to determine if observed results are rare enough to justifying rejecting H_0. For example, if $\alpha = 0.05$, we are willing to be wrong 5% of the time. We are given $\sigma = 2, n = 30$. To begin the test, we assume H_0 is true:

$$P(\bar{x} < \mu_{\bar{x}}) = P(Z < \frac{\mu_{\bar{x}} - \mu}{\sigma_{\bar{x}}})$$

$$P(\bar{x} < 44.5) = P(Z > \frac{44.5 - 45}{0.365}) = P(Z < -1.37) = 0.0835$$

To conclude, we compare our answer with α. If the probability is greater than the significance level we accept H_0, else we accept H_1. In this case $0.0835 > \alpha$, therefore, the evidence isn't sufficient to refute our claim.