Normal probability distribution.

This is one of the most important continuous distributions.

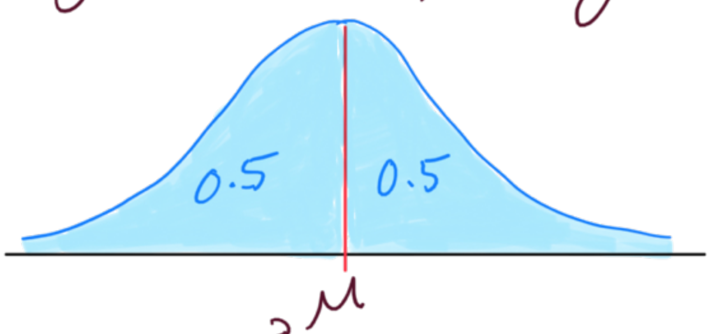
For example, the following variables and many others have probability distributions:-

a) The weights or heights of humans.

b) The blood pressure of humans.

c) The time taken to perform a computer alogorithm.

The curve for a typical normal probability distribution looks like:-



mean = median = mode.

Notation: We use $X \sim N(\mu, \sigma^2)$ to mean that the random variable X has a normal distribution with mean μ and variance σ^2.

Fact: Let $X \sim N(\mu, \sigma^2)$, then the p.d.f. of $x \in \mathbb{R}$ is given by $\boxed{f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}}$.

or continuous.

The probability of the normal distribution.

The following results are not only true for the normal distribution, but also for any continuous distribution.

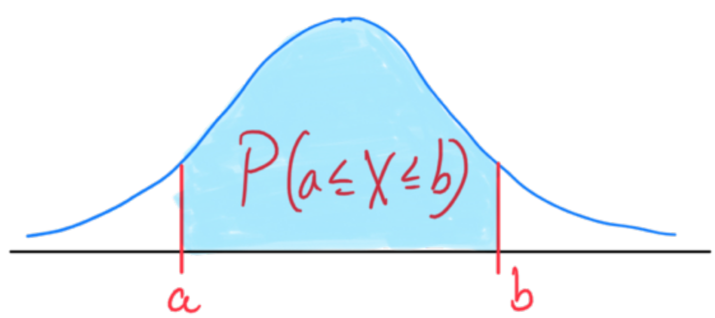
Fact: Let $X \sim N(\mu, \sigma^2)$ with p.d.f. $f(x)$, then

1) the probability that X assumes a value in a set $A \subset \mathbb{R}$ is given by $P(X \in A) = \int_{x \in A} f(x)\,dx$.

2) As a special case of (1), we have

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx \ , \quad \text{assuming } a < b.$$

$$P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$



3) As a special case of (2), we have

$$P(X=a) = \int_a^a f(x)\,dx = 0.$$

Linearity property.

If $X \sim N(\mu, \sigma^2)$, then $aX+b \sim N(a\mu+b, a^2\sigma^2)$.

Here a and b are any real numbers.

Ex. If $X \sim N(20,16)$ and $Y = 3X+4$, then

$E(Y) = E(3X+4) = 3E(X)+4 = 3(20)+4 = 64.$

$Var(Y) = 3^2(E(X))^2 = 3^2(16) = 144.$

Thus, $Y \sim N(64, 144)$.

Standard normal distribution.

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X + \left(\frac{-\mu}{\sigma}\right)$

is normally distributed with

$E(Z) = \frac{1}{\sigma}\mu + \left(\frac{-\mu}{\sigma}\right) = 0$ and $Var(\sigma) = \left(\frac{1}{\sigma^2}\right)\sigma^2 = 1.$

Thus, $Z \sim N(1,0)$.

Def. The $N(1,0)$ distribution is called the standard normal distribution.

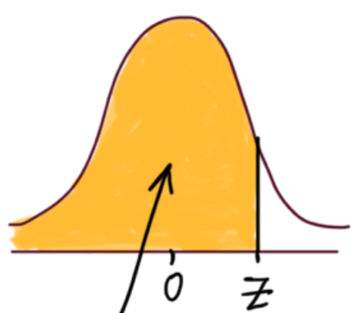
The normal distribution tables (the Z-tables)

The normal probability tables give the commulative probability $P(Z \le z)$ for all different values of z.

Area under the normal curve

z	$*$	$*$	d_3
$*$	$*$	$*$	$*$
$*$	$*$	$*$	$*$
$d_1 . d_2$	$*$	$*$	$T(d_1 . d_2 d_3)$

$P(Z \le d_1 . d_2 d_3)$



$P(Z \le z)$

z	0.00	0.01	0.02	0.03	0.04
1.0	*	*	*	0.8485	*
1.1	*	*	*	0.8708	*
1.2	*	*	*	0.8907	*
1.3	*	*	*	0.9082	*
1.4	*	*	*	0.9236	*
1.5	*	*	*	0.9370	*
1.6	*	*	*	0.9484	*
1.7	*	*	*	0.9582	*
1.8	*	*	*	0.9664	*
1.9	*	*	*	0.9732	*

<u>Ex</u>. If $Z \sim N(0,1)$, then

i) $P(Z \leq 1.13) = T(1.13) = 0.8708$

ii) $P(Z < 1.13) = P(Z \leq 1.13) = 0.8708.$

iii) $P(Z > 1.13) = 1 - P(Z \leq 1.13) = 1 - 0.8708 = 0.1292.$

iv) $P(1.13 \leq Z \leq 1.83) = P(Z \leq 1.83) - P(Z < 1.13)$

$= 0.9664 - 0.8708 = 0.0957.$

<u>Ex</u>. If $Z \sim N(0,1)$, find the value of

i) r such that $P(Z \leq r) = 0.9370$.

ii) s such that $P(Z \geq s) = 0.0918$.

ii) t such that $P(t < Z < 0.51) = 0.4$.

<u>Soln.</u> i) From the table, it is clear that $r = 1.53$.

ii) $P(Z \geq s) = 0.0918$ is equivalent to

$P(Z \leq s) = 1 - 0.0918 = \underline{0.9082}$.

From the table, $s = 1.33$.

iii) $P(t < Z < 0.51) = 0.4$

$P(Z < 0.51) - P(Z < t) = 0.4$

$0.6950 - P(Z < t) = 0.4$ (from tables)

$P(Z < t) = 0.6950 - 0.4 = 0.2950$.

From tables, we have $t = -0.5 + 0.04 = -0.54$.

<u>Remark</u>: We can find probabilities about any normal variables (not only the standard variable Z) by chaning to the standard normal distribution.

<u>Ex.</u> If $X \sim N(20, 16)$, find

i) $P(X \leq 24.12)$.

ii) $P(X > 25.72)$

iii) $P(25.72 \leq X \leq 27.72)$

Soln. Note that $\mu = E(X) = 20$ and $\sigma = 4$.

i) $P(X \leq 24.12) = P\left(\frac{X-\mu}{\sigma} \leq \frac{24.12-20}{4}\right)$

$= P(Z \leq 1.03)$

$= 0.8485$.

ii) $P(X > 25.72) = 1 - P(X \leq 25.72)$

$= 1 - P\left(\frac{X-\mu}{\sigma} \leq \frac{25.72-20}{4}\right)$

$= 1 - P(Z \leq 1.43)$

$= 1 - 0.9236$

$= 0.0764$.

iii) $P(25.72 \leq X \leq 27.72) = \underline{P(X \leq 27.72)} - \underline{P(X < 25.72)}$

Now, $P(X \leq 27.72) = P\left(\frac{X-\mu}{\sigma} \leq \frac{27.72-20}{4}\right)$

$= P(Z \leq 1.93)$

$= 0.9732$.

$P(X < 25.72) = 1 - P(X \geq 25.72)$

$= 1 - 0.0764$

$= 0.9236$.

Thus, $P(25.72 \leq X \leq 27.72) = 0.9732 - 0.9236 = 0.0496$.

Ex:- The grades of Math 131 has $N(\overset{\mu}{68}, \overset{\sigma^2}{100})$ distribution.

Find a) the proportion of grades that are more than 85.

b) the proportion of grades between 60 and 90.

c) the 95th-percentile.

Soln. a) $P(X>85) = P\left(\frac{X-\mu}{\sigma} > \frac{85-68}{10}\right)$

$= P(Z > 1.7)$

$= 1 - T(1.7)$

$= 1 - 0.9554$ (from tables).

$= 0.0446$

b) $P(60 < X < 90) = P\left(\frac{60-68}{10} < \frac{X-\mu}{\sigma} < \frac{90-68}{10}\right)$

$= P(-0.8 < Z < 2.2)$

$= T(2.2) - T(-0.8)$

$= 0.9861 - 0.2119$ (from tables)

$= 0.7742$

c) Let P_{95} be the 95th-percentile, then

$P(X \le P_{95}) = 0.95$ or $P\left(\frac{X-\mu}{\sigma} \le \frac{P_{95}-68}{10}\right) = 0.95$

$P\left(Z \le \frac{P_{95}-68}{10}\right) = 0.95$

From the Z-table, we have $\frac{P_{95} - 68}{10} = 1.64$.

This implies that $P_{95} = 84.4$.

So, if 10% of the students will get A, then the least grade that will get A is 84.4.

<u>Another solution</u>; We use the rule:-

$$\boxed{P_{100p} \text{ of } X = \left(P_{100p} \text{ of } Z\right)\sigma_X + \mu_X}.$$

So, P_{95} of $X = (1.64)(10) + 68 = 84.4$.

The normal approximation to the binomial distribution.

We might use the normal distribution as an approximation for the binomial distribution.

Let $X \sim B(n, p)$, (here n is large and p is moderate, $np \geq 5$ and $n(1-p) \geq 5$) then this binomial distribution can be approximated by the normal distribution $N(\underbrace{np}_{\mu}, \underbrace{np(1-p)}_{\sigma^2})$.

Because the binomial distribution is discrete while the normal is continuous, we change the discrete values of the binomial distribution to continuous one by the so called <u>continuity correction</u>:

$X = a$ becomes $X \in (a-\frac{1}{2}, a+\frac{1}{2})$. So

- $P(X=a)$ is corrected to $P(a-\frac{1}{2} < X < a+\frac{1}{2})$.
- $P(a \le X \le b)$ " " " $P(a-\frac{1}{2} < X < b+\frac{1}{2})$.
- $P(a < X < b)$ " " " $P(a+\frac{1}{2} < X < b-\frac{1}{2})$.
- $P(a < X \le b)$ " " " $P(a+\frac{1}{2} < X < b+\frac{1}{2})$.
- $P(a \le X < b)$ " " " $P(a-\frac{1}{2} < X < b-\frac{1}{2})$.

<u>Ex</u>. Let $X \sim B(10, 0.5)$. Find $P(X \in \{2,3,4\})$ using
a) the binomial tables b) the normal approximation.

<u>Soln</u>. a)
$$P(X \in \{2,3,4\}) = P(X \le 4) - P(X \le 1)$$
$$= T(4) - T(1)$$
$$= 0.377 - 0.011$$
$$= 0.366.$$

b) $\mu = np = 5$ and $\sigma^2 = np(1-p) = 2.5$.

We will use $N(5, 2.5)$ instead of $B(10, 0.5)$.

$$P(X \in \{2,3,4\}) = P(2 \leq X \leq 4)$$
$$\approx P(1.5 < X < 4.5) \quad \text{(Continuity Correction)}$$
$$= P\left(\frac{1.5-5}{\sqrt{2.5}} \leq \frac{X-\mu}{\sigma} \leq \frac{4.5-5}{\sqrt{2.5}}\right)$$
$$= P(-2.22 \leq Z \leq -0.32)$$
$$= T(-0.32) - T(-2.22)$$
$$= 0.3745 - 0.0132$$
$$= 0.3613.$$

Ex. Let $X \sim B(100, 0.2)$. Use the normal approximation to find approximate values for the probabilities:

a) $P(X < 26)$.

b) $P(18 < X < 26)$.

c) $P(18 \leq X < 26)$.

d) $P(18 \leq X \leq 26)$.

e) $P(X = 26)$.

f) $P(X > 26)$.

<u>Soln.</u> Because $X \sim B(100, 0.2)$, we have

$\mu = np = 20$ and $\sigma^2 = np(1-p) = 16$.

We will use $N(20, 16)$ instead of $B(100, 0.2)$.

a) $P(X < 26) \approx P(X \leq 25.5)$

$$= P\left(\frac{X-\mu}{\sigma} \leq \frac{25.5-20}{4}\right)$$

$$= P(Z \leq 1.38)$$

$$= 0.9162.$$

b) $P(18 < X \leq 26) \approx P(18.5 < X < 26.5)$

$$= P\left(\frac{18.5-20}{4} < \frac{X-\mu}{\sigma} < \frac{26.5-20}{4}\right)$$

$$= P(-0.38 < Z < 1.63)$$

$$= T(1.63) - T(-0.38)$$

$$= 0.9484 - 0.3520$$

$$= 0.5964.$$

c) $P(18 \leq X < 26) \approx P(17.5 < X < 25.5)$

$$= \overset{\text{Exc}}{\cdots} = 0.6519$$

d) $P(18 \leq X \leq 26) \approx P(17.5 < X < 26.5)$

$$= \overset{\text{Exc}}{\cdots} = 0.6841.$$

e) $P(X = 26) \approx P(25.5 < X < 26.5)$

$$= \cdots \quad \text{Exc.}$$

f) $P(X > 26) = P(X \geq 27)$

$$\approx P(X > 26.5)$$

$$= P\left(\frac{X-\mu}{\sigma} > \frac{26.5-20}{4}\right)$$

$$= P(Z > 1.65)$$

$$= 1 - P(Z \leq 1.63)$$

$$= 1 - 0.9484$$

$$= 0.0516.$$

This is Example 4.2.2 page 130 in the book.

Ex. Assume that 10% of heavy smokers will suffer from lung cancer after the age of 40. In a sample of 100 heavy smokers, what is the probability that

a) at least 12 will have lung cancer?

b) not more that 14 will have lung cancer?

c) exactly 12 will have lung cancer?

Soln. Let X be the number of those that will have lung cancer out of 100. Then $X \sim B(100, 0.1)$.

Hence $\mu = np = 10$ and $\sigma = \sqrt{np(1-p)} = \sqrt{9} = 3$.

We will use $N(10, 9)$ instead of $B(100, 0.1)$.

a) $P(X \geq 12) = 1 - P(X < 12)$

$$\approx 1 - P(X < 11.5)$$

$$= \overset{\text{Exc}}{\dashrightarrow} = 0.30915.$$

b) $P(X \leq 14) \approx P(X < 14.5)$

$$= \overset{\text{Exc}}{\dashrightarrow} = 0.9332.$$

c) $P(X=2) \approx P(11.5 < X < 12.5)$

$$= \overset{\text{Exc}}{\cdots} = 0.1052.$$

The central limit theorem (C.L.T.)

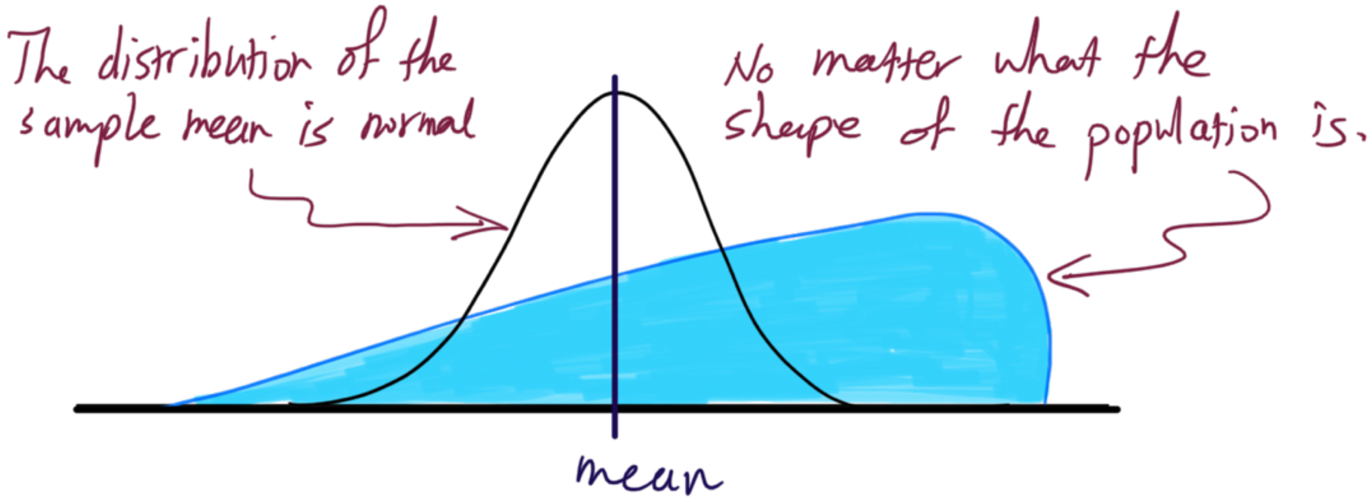
The C.L.T. is very fundamental in probability theory.

Suppose we have a random variable X with expected value $E(X) = \mu$ and variance σ^2.

We extract a random sample $X_1, X_2, \ldots, X_n$ from X with replacement, then the sample mean $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ distributes with the expected value μ and variance σ^2/n.

In case $n \longrightarrow \infty$ (in practice $n \geq 30$)

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, whatever the distribution of X be.



Furthermore, if X is normally distributed, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ even if n is small.

That is, if $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ even if $n < 30$.

Ex. Let $\bar{X}$ be the mean of a sample of size 100, randomly selected from a population that has mean $\mu = 20$ and variance $\sigma^2 = 400$. Find

a) the distribution of $\bar{X}$.

b) $P(\bar{X} > 22)$.

c) the 95^{th} percentile of $\bar{X}$.

Soln a) Note that $n = 100 > 30$, so $\bar{X} \sim N\left(20, \frac{400}{100}\right)$.

Thus, $\bar{X} \sim N(20, 4)$.

b) $$P(\bar{X} > 22) = P\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{22 - 20}{\sqrt{4}}\right)$$

$$= P(Z > 1)$$

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

$$= 1 - 0.8413$$

$$= 0.1587.$$

c) We use the rule :-

$$\boxed{P_{100p} \text{ of } \bar{X} = \left(P_{100p} \text{ of } Z\right) \sigma_{\bar{X}} + \mu_{\bar{X}}}.$$

So $$P_{95} \text{ of } \bar{X} = \left(P_{95} \text{ of } N(20, 4)\right) * 2 + 20$$

$$= 1.64 * 2 + 20$$

$$= 23.29.$$

Exc. (A) The weight of random group of children has $N(3.5, 1.2)$.

i) If a newborn is randomly selected, find the probability that his/her weight is greater than 3 kg.

ii) Find the 80^{th} percentile of these weights.

(B) Let $\bar{X}$ be the mean weight of a sample of size 16 children that are randomly selected from those in part (A).

i) Find $P(\bar{X} > 3)$.

ii) Find the 80^{th} percentile of $\bar{X}$.

Final answer: (A) i) 0.6772

ii) 4.42

(B) i) 0.9664

ii) 3.73

Using C.L.T.

Exc.

Searching keywords:

- Normal distribution
- Normal approximation to the binomial distribution
- The central limit theorem
- Find the probability of
- The University of Jordan الجامعة الأردنية
- Principles of Statistics مبادئ الإحصاء
- Baha Alzalg بهاء الزالق

References: See the course website

http://sites.ju.edu.jo/sites/Alzalg/Pages/131.aspx

For any comments or concerns, please use my email to contact me.

د. بهاء محمود الزالق
The University of Jordan
Dr. Baha Alzalg
baha2math@gmail.com