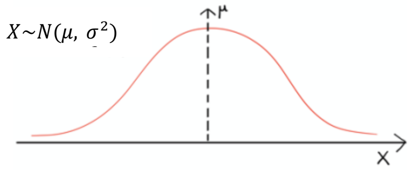
Normal Distribution Cheat Sheet

Stats/Mech Year 2

The normal distribution is a continuous probability distribution that can be used to model a vast number of naturally occurring scenarios. As a result, it is one of the most important probability distributions in statistics. Examples of natural variables that follow the normal distribution are IQ scores, height and weight. The key difference between the normal distribution and the distributions you have previously met in Year 1 is that the normal distribution is continuous, while the others are discrete.

Characteristics

The normal distribution:



- has two parameters: the mean, μ, and variance σ^2.
- Is symmetrical (mean = median = mode).
- has a bell-shaped curve with asymptotes at each end.
- has total area under the curve equal to 1.
- has $P(X=a)=0$ for any a. This is true for any continuous distribution.

If a variable X follows a normal distribution with mean μ and variance σ^2, we write $X\sim N(\mu,\sigma^2)$.
It is often very helpful to sketch the normal curve when solving normal distribution problems.

Finding probabilities

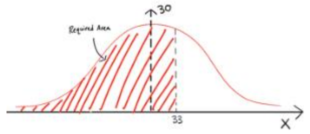
- You need to be able to use the normal cumulative distribution function on your calculator to find probabilities.

You will need to enter a lower and upper bound when using this function, as well as the mean and standard deviation of the distribution being used.

Example 1: The random variable $X\sim N(30,2^2)$. Find $P(X<33)$.

Using the cumulative function on your calculator, we enter our mean as 30 and our standard deviation as 2. Our upper bound will be 33 and for our lower bound we need to enter a really small value. Remember that the normal curve is asymptotic at both ends so to find an accurate approximation for the total area to the left of $x=33$ we take our lower bound to be a small value. Take for example, -500.

$\Rightarrow P(X<33)=0.933\ (3\ s.f)$



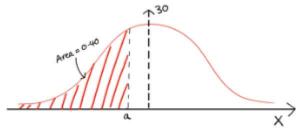
The inverse normal distribution function

- You can use the inverse normal function on your calculator to find the value of a such that $P(X<a)=p$.

Some problems might require you to instead find the value of a such that $P(X>a)=p$. Be aware that most calculators (e.g. the Casio fx-991ex) will only return the value of a such that $P(X<a)=p$, so you will need to use the property $P(X>a)=1-P(X<a)$ in such situations. See example 3 for more detail.

Example 2: Given that $X\sim N(30,2^2)$, find the value of a such that $P(X<a)=0.4$.

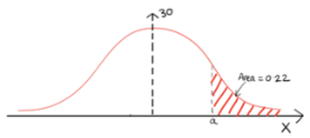
Using the inverse normal function with $p=0.4$, mean = 30 and standard deviation = 2 gives us $a=29.5$.



Example 3: Given that $X\sim N(30,2^2)$, find the value of a such that $P(X>a)=0.22$.

We first manipulate our expression:

$P(X>a)=1-P(X<a)=0.22$
$\therefore P(X<a)=0.78$



Now using the inverse function with $p=0.78, \mu=30$ and $\sigma=2$ gives us that $a=31.5$.

Note that some calculators (e.g. the Casio CG-50) will allow you to specify which tail of the distribution the probability p corresponds to, eliminating the need for any manipulation for problems similar to example 3.

The standard normal distribution

We can use a coding to standardise our data, making it much easier to analyse and work with. For problems where the mean and/or variance are unknown, it is very useful to code data via the standard normal distribution.

- The standard normal distribution denoted $Z\sim N(0.1^2)$, has mean 0 and standard deviation 1.
- If $X\sim N(\mu,\sigma^2)$, you can use the coding $Z=\frac{X-\mu}{\sigma}$ to convert your variable into a standard normal variable.
- The notation $\Phi(a)$ is equivalent to $P(Z<a)$.

Example 4: The random variable $X\sim N(50,4^2)$. Write $P(X\geq 55)$ in terms of $\Phi(z)$ for some value z.

Since the normal distribution is continuous: $P(X\geq 55)=P(X>55)$

$P(X>55)=1-P(X<55)$

Converting to standard normal:

$$P(X<55)=P\left(Z<\frac{55-\mu}{\sigma}\right)=P\left(Z<\frac{55-50}{4}\right)=P(Z<1.25)$$

So $P(X\geq 55)=1-P(Z<1.25)=1-\Phi(1.25)$

Finding μ and σ

You need to be able to use the standard normal distribution to solve problems where you must find the mean and/or variance. You will be given either one or two probabilities which you must standardise in order to find the unknown parameters. We will go through two examples; one in which one parameter is missing and the other where both parameters are missing.

Example 5: The random variable $X\sim N(\mu,5^2)$ and $P(X<18)=0.9032$. Find the value of μ.

We are told that $P(X<18)=0.9032$. Standardising:

$P(X<18)=P\left(Z<\frac{18-\mu}{5}\right)=0.9032$

We can now use the inverse normal function (with $\mu=0, \sigma=1, area=0.9032$) to find the value of a such that $P(X<a)=0.9032$:

The calculator returns a value of $a=1.30$

$\therefore \frac{18-\mu}{5}=1.30 \Rightarrow\ \mu=11.5.$

Example 6: The random variable $X\sim N(\mu,\sigma^2)$. Given that $P(X<17)=0.8159$ and $P(X<25)=0.9970$, find the value of μ and σ.

We are told $(X<17)=0.8159$ and $P(X<25)=0.9970$. We standardise both of these cases separately, following the same method as in example 4:

$\Rightarrow P(X<17)=P\left(Z<\frac{17-\mu}{\sigma}\right)=0.8159$
$\Rightarrow P(X<25)=P\left(Z<\frac{25-\mu}{\sigma}\right)=0.9970$

Using the inverse normal function for both probabilities, we acquire two equations:

$\frac{17-\mu}{\sigma}=-0.8998 \Rightarrow 17-\mu=-0.893\sigma$

$\frac{25-\mu}{\sigma}=2.748 \Rightarrow 25-\mu=2.748\sigma$

We now have two equations with two unknowns, so we can solve for μ and σ. You could use your calculator for this part.

Solving gives us $\mu=19.0\ ,\sigma=2.20\ (3\ s.f)$

Approximating a binomial distribution

Under certain conditions, the binomial distribution is statistically very similar to the normal distribution. As a result, the normal distribution can be used as an approximation for the binomial distribution.

- If n is large and p is close to 0.5, the binomial distribution $X\sim B[n,p]$ can be approximated by the normal distribution $N(\mu,\sigma^2)$, where
  - $\mu=np$
  - $\sigma=np(1-p)$

Recall that the mean and the variance of a binomially distributed random variable are np and $np(1-p)$ respectively.

The reason that p has to be close to 0.5 is because the normal distribution is symmetrical. Note that there is no specific range that p must fall into for this approximation to be valid. The closer p is to 0.5, the better the approximation is

However, there is a small problem: the binomial distribution is discrete while normal is continuous. This raises an issue of inaccuracy with this approximation. To see why, take for example the probability $P(X>5)$ where X is discrete. This probability is associated with X taking the values 6, 7, 8, 9 and beyond. If we use a normal approximation and find $P(X>5)$ then we are finding the probability X is between 5 and 6 as well as anything beyond, which is inaccurate since we didn't want any values of X between 5 and 6 to begin with. To tackle this issue, we can apply a process known as the continuity correction:

- When using a normal approximation to approximate a binomial distribution, you must apply a continuity correction to ensure maximal accuracy when calculating probabilities.

Applying the continuity correction is as simple as adding or subtracting 0.5 from your value. You can use the following table to help you decide when to add or subtract 0.5:

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

Example 7: A drill bit manufacturer claims that 52% of its bits last longer than 40 hours. A random sample of 600 bits is taken. Using a suitable approximation, find the probability between 300 and 350 bits last longer than 40 hours.

Let X be the number of bits that last longer than 40 hours in a batch of 600. Then $X\sim B[600,0.52]$. Since n is large and p is close to 0.5, we can use a normal approximation with $\mu=np,\ \sigma=np(1-p)$

$\Rightarrow np=600(0.52)=312$
$\Rightarrow np(1-p)=149.76$

$\therefore X\approx\sim N(312,149.76)$

We want to find $P(300<X<350)$. Applying the continuity correction, this becomes: $P(300.5<X<349.5)$. Using the calculator to find this probability gives us an answer of 0.846.

Hypothesis testing with the normal distribution

You also need to be able to test hypothesis regarding the mean of a normal distribution, by looking at a sample. You will need to use the following fact:

- For a random sample of size n taken from a random variable $X\sim N(\mu,\sigma^2)$, the sample mean $\overline{X}$ is normally distributed with $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$.

The idea is that we use the sample mean distribution to see whether the mean from an actual sample is significant enough to reject the null hypothesis.

Example 8: The diameters of circular cardboard drink mats produced by a particular machine are normally distributed with mean 9cm and standard deviation 0.15cm. After the machine is serviced a random sample of 30 mats is selected and their diameters are measured to see if the mean diameter has decreased. The mean of the sample was 8.95cm. Test, at the 5% significance level, if there is significant evidence to suggest the mean diameter of the machine has decreased.

We are testing whether the mean has decreased, so this is a one-tailed test.
$\therefore$ our hypotheses are: $H_0: \mu=9$, $H_1: \mu\neq 9$

Let X be the diameter of the drink mats. Then $X\sim N(9,0.15^2)$.
Our sample mean distribution will therefore be $\overline{X}\sim N(9,\frac{0.15^2}{30})$ since the sample size is 30.
We use this to find $P(\overline{X}<8.95)$ and compare this value to 2.5%.

Using a calculator, we find that $P(\overline{X}<8.95)=0.034>0.025$

$\therefore$ Our result is insignificant and we can conclude that there is insufficient evidence to suggest the mean has decreased (accept H_0).