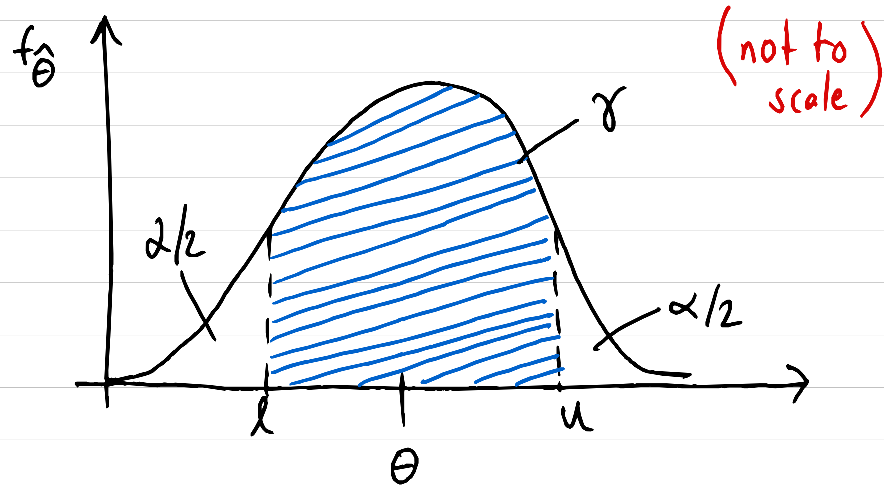
23) Confidence intervals for the mean

Consider the distribution of an estimator $\hat{\Theta}$ for some model parameter θ. If the estimator is at all good then its density will be concentrated near the true value θ.

In this chapter we assume that the estimator is a continuous random variable.

The following figure sketches an example of a density function



We know however that, the probability that the estimator will give exactly the correct result is 0 as by Thm 5.4 for any continuous random variable,

$$\boxed{P(\hat{\Theta}=\theta)=0}$$

The best we can hope for is that with high probability the estimator gives a value close to the true value, say within an interval from $\theta - a$ to $\theta + b$ for some $a, b \in \mathbb{R}$

So we consider the probability

$$P(\theta - a \leq \hat{\Theta} \leq \theta + b) = \gamma$$

$$\Rightarrow \quad P(\theta - a \leq \hat{\Theta} \leq \theta + b) = 1 - \alpha$$

The probability γ is the area is the area of the shaded region under density function in the figure drawn previously.

In the special case where the density function is symmetric around $x = \theta$ and $a = b$, the remaining probability $\alpha = 1 - \gamma$, is split equally between right and left tail again as drawn previously.

The equations above are not very useful to us yet, because we do not know the true value of θ and therefore do not know location of the interval $(\theta - a, \theta + b)$.

However we can use that

$$P(\theta - a < \hat{\Theta} < \theta + b)$$

$$= \boxed{P(\hat{\Theta} - b < \theta < \hat{\Theta} + a) = \gamma}$$

Now we have a random interval

$$\boxed{(L, U) = (\hat{\Theta} - b,\ \hat{\Theta} + a)}$$

that contains the true value with probability γ.

When we evaluate the random variables L & U on our data we obtain the so-called confidence intervals.

The random variables L and U give the lower and upper end of random interval are not always in the form given above.

The defn given follows a more general case

Defn 23.1: Suppose a dataset $x_1, \ldots, x_n$ is modelled by random variables $X_1, \ldots, X_n$. Let θ be the parameter and $\gamma \in [0, 1]$.

If there exists random variables

$$\underline{L = g(X_1, \ldots, X_n)} \quad \text{and} \quad \underline{U = h(X_1, \ldots, X_n)}$$

such that

$$\boxed{P(L < \theta < U) = \gamma}$$

for any value of θ. Then the interval

$$\boxed{(l, u)}$$

is a $100\gamma\%$ confindence interval for θ where

$$\underline{l = g(x_1, \ldots, x_n)} \quad \text{and} \quad \underline{u = h(x_1, \ldots, x_n)}$$

γ is the confidence level. If we only have

$$\boxed{P(L \leq \theta \leq U) \geq \gamma}$$

then we only speak of a conservative confidence interval.

Note that while the random interval (L, U) contains the true value θ with probability γ, it would be incorrect to say that therefore the interval (l, u) contains the true value θ with probability γ.

Once we have evaluated the random variables using the data, a traditional statistician would no longer speak of probability. We now have a definite interval and the true value either lies in it or does not.

It is the same as when your football team has played, the game is over, but you are away and have not yet heard the result. Even though you don't know the result yet, your team has either won or they have lost. There is nothing you can do about it anymore.

You can still speak about how confidently you believe that they have won, but you should not speak of the probability that they have won.

Hence we call γ the confidence level.

In this module we will concentrate on the case where the data is modelled as an iid sample and the model parameter for which we want to know the confidence interval is the expectation of the model distribution.

Let us first consider the case where i.i.d sample is from a normal distribution.

Notation: Recall standard normal distribution.

$$Z \sim N(0,1) \; ; \; F_Z = P(Z \leq z) = \Phi(z).$$

$$f_Z(z) = \phi(z)$$

let z_p be the percentile. By Chapter 5, quantile is found by the inverse function of distribution function

Let

$$z_p = \Phi^{-1}(1-p) = (1-p) \text{ quantile}$$

- Quantile function: it gives us the value such that it gives us a certain probability to the left.

Quantile function input: probability $p \in [0,1]$

Quantile function output: a real number $q \in \mathbb{R}$
such that
$P(X \leq q) = p$

Input probability p $\rightarrow$ output real number $q \in \mathbb{R}$
such that
$P(X \leq q) = p = F_X(q)$

So specifically for standard normal:

Quantile function is

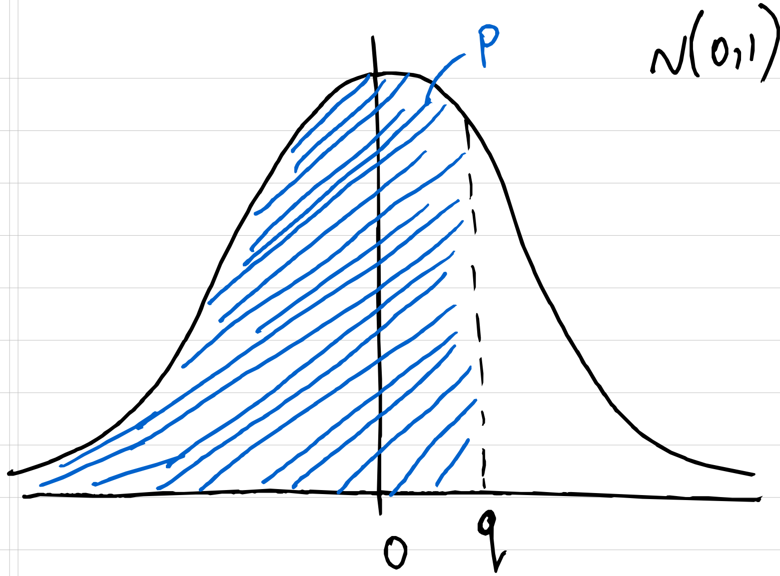
Input probability p $\longrightarrow$ output real number $q \in \mathbb{R}$
such that
$P(X \leq q) = \Phi(q) = p$

So we can say that quantile function is

$$F_X^{-1} : [0,1] \rightarrow \mathbb{R} \; ; \; F_X^{-1}(p) = q$$

For standard normal

$$\Phi^{-1} : [0,1] \rightarrow \mathbb{R} \; ; \; \Phi^{-1}(p) = q$$



$$P(X \leq q) = F_X(q) = p \Rightarrow F_X^{-1}(p) = q$$

For standard normal:

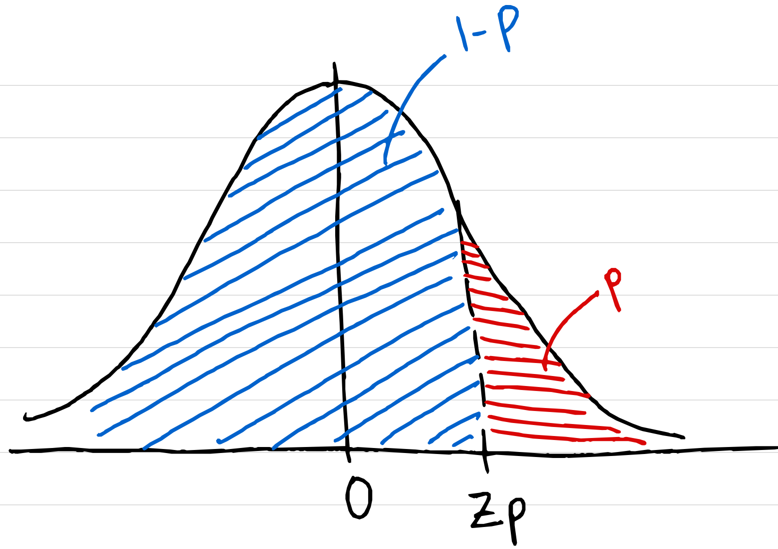
$$\Phi^{-1}(0.5) = 0.5 \text{ quantile} = 0 = q$$

$$\Phi^{-1}(0.84) = 0.84 \text{ quantile} \approx 1 = q$$

Quantile of $(1-p)$ is $\Phi^{-1}(1-p)$ is the number z_p such that

$$\Phi(z_p) = P(Z \leq z_p) = 1-p$$

(diagram on next page)



So

$$P(Z \le z_p) = 1-p \Rightarrow 1 - P(Z \ge z_p) = 1-p$$

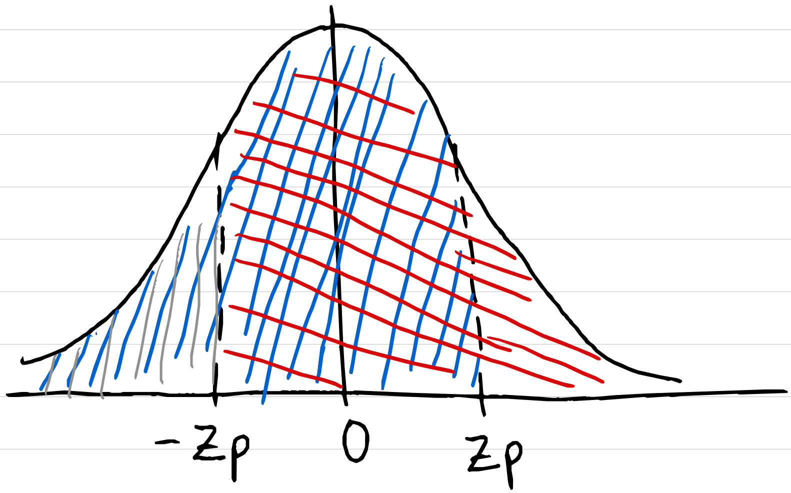
$$\Rightarrow P(Z \ge z_p) = p$$

By defn of standard normal

$$\begin{aligned} P(Z \ge z_p) &= 1 - P(Z \le z_p) \\ &= 1 - \Phi(z_p) \\ &= 1-(1-p) \\ &= p \end{aligned}$$

$$\Rightarrow \quad P(Z \ge z_p) = p$$

Because of symmetry of standard normal distribution:



$\Phi(z_p) = 1 - \Phi(-z_p)$ (proven in ws 5)

We also have

$$\Phi(z_p) = 1 - \Phi(-z_p) \Rightarrow \Phi(-z_p) = 1 - \Phi(z_p)$$

$$\Rightarrow \Phi(-z_p) = 1 - (1-p)$$

$$\Rightarrow \Phi(-z_p) = p$$

So

$$P(Z \geq -z_p) = 1 - P(Z \leq -z_p)$$
$$= 1 - \Phi(-z_p)$$
$$= 1 - p$$

Also since

$\Phi(z_p) = 1 - p$, by changing subscripts;

$$\Phi(z_{1-p}) = 1 - (1-p) = p$$

$\Rightarrow$

$$\Phi(z_{1-p}) = p$$

So

$$\Phi(-z_p) = \Phi(z_{1-p})$$

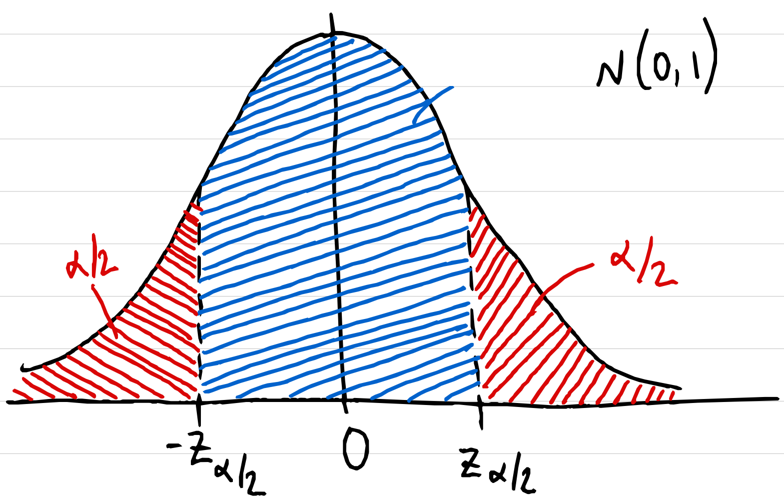
$$\Phi(-z_p) = \Phi(z_{1-p})$$

$$\Rightarrow \Phi^{-1}(\Phi(-z_p)) = \Phi^{-1}(\Phi(z_{1-p}))$$

$$\Rightarrow \underline{-z_p = z_{1-p}}$$

[$f^{-1}(f(x)) = x$ and $f(f^{-1}(x)) = x$
See maths skills 1]

Now from the following diagram



Now we need to evaluate

$$P(-z_{\alpha/2} < Z < z_{\alpha/2})$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = P(Z < z_{\alpha/2}) - P(Z < -z_{\alpha/2})$$

$$= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})$$

$$= \Phi(z_{\alpha/2}) - \Phi(z_{1-\alpha/2})$$

(as established; $-z_p = z_{1-p}$ with $p = \alpha/2$)

$$= 1 - \alpha/2 - \alpha/2$$

(as established before; $\Phi(z_p) = 1-p$, $\Phi(z_{1-p}) = p$)

$$= 1 - \alpha$$

$\Rightarrow$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

Theorem 23.2: Suppose a dataset $x_1, \dots, x_n$ is modelled as an iid sample $X_1, \dots, X_n$ from an $N(\mu, \sigma^2)$ distribution with unknown mean but known variance. Then the interval

$$\boxed{\left(\bar{x}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} , \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)}$$

is a $100(1-\alpha)\%$ confidence interval for mean μ.

proof: According to defn 23.1, we need to show that

$$P\left(\bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1-\alpha$$

We know that the sample mean

$$\bar{X}_n = \left(\frac{X_1 + \dots + X_n}{n} \right)$$

is normally distributed

$$\bar{X}_n \sim N(\mu, \sigma^2/n)$$

Therefore

$$Z = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Therefore

$$1-\alpha = P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right)$$

$$= P\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} < z_{\alpha/2}\right)$$

$$= P\left(-\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

(multiplying inequality by -1)

$$= P\left(\bar{X}_n - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$$

$\Rightarrow$

$$P\left(\bar{X}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$\blacksquare$