Symmetric distribution. Z is distributed symmetrically about 0 if Z and $-Z$ have the same distribution. If Z has density f_Z, then note, in the presence of symmetry, $f_Z(z) = f_Z(-z)$ for all z. X is symmetric about μ if $X - \mu$ is symmetric about 0, or $X - \mu$ and $-(X - \mu)$ have the same distribution. If X has density f_X, then we need $f_X(\mu + x) = f_X(\mu - x)$ for all x. ($f_X(\mu + x) = f_Z(\mu + x - \mu) = f_Z(x) = f_Z(-x)$; $f_X(\mu - x) = f_Z(\mu - x - \mu) = f_Z(-x) = f_Z(x)$)

A statistical model can arise in two different ways, and hence may have two different interpretations.
(i) Measurement error model: Suppose we want to determine the length μ of a table by measuring it using a tape measure. Then any measurement X can be represented as $X = \mu + \epsilon$, where ϵ stands for the deviation from the true length μ due to measurement error. If $\epsilon \sim N(0, \sigma^2)$, then $X \sim N(\mu, \sigma^2)$. σ provides a measure of how large a typical deviation from μ can be.
(ii) Sampling from a population: Consider collecting a random sample (independent and identically distributed or i.i.d.) of observations from a population to determine certain features such as height, weight or family income. Now there is variation within the population. Therefore, if we model an observation X as $X \sim N(\mu, \sigma^2)$, then μ represents the population average and σ measures the deviation or spread of the population around μ.

The statistical inferential procedure will be the same, irrespective of how the model is arrived at. i.e., it is for μ and σ^2 of $N(\mu, \sigma^2)$ whether the observation came from (i) or (ii).

Location-Scale Families Consider $U \sim U(-1, 1)$ with density $f_U(u) = \frac{1}{2} I_{(-1,1)}(u)$, and let $X = \mu + U$. Then $X \sim U(\mu - 1, \mu + 1)$ with density $f_X(x) = \frac{1}{2} I_{(\mu-1,\mu+1)}(x) = \frac{1}{2} I_{(-1,1)}(x - \mu) = f_Z(x - \mu)$. In other words, the location of X is a translation by μ of the location of U. The family of distributions for X indexed by μ is called a location family with location parameter μ. Note that μ is location for X if $X - \mu$ has a distribution which is free of μ.

Similarly, if $Z \sim N(0, 1)$ with density $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$, then $X = \sigma Z$, $\sigma > 0$, then $X \sim N(0, \sigma^2)$ with density $f_X(x) =$
$\frac{1}{\sqrt{2\pi}\sigma} \exp(-(x - \mu)^2/(2\sigma^2)) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp(-((x - \mu)/\sigma)^2/2) = \frac{1}{\sigma} f_Z(\frac{x-\mu}{\sigma})$.
In this case, X is scaled by σ, and the family of distributions for X indexed by σ is called a scale family with scale parameter σ. It is important to note that σ is scale for X if X/σ has a distribution which is free of σ. Combining location and scale gives the location-scale family.

Definition Let X be a real-valued random variable, with density

$$f(x|\mu,\sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right),$$

where g is also a density function, $-\infty < \mu < \infty$, $\sigma > 0$. The parameters μ and σ are called location and scale parameters.

With X as above, $Z = (X-\mu)/\sigma$ has density g. The normal $N(\mu, \sigma^2)$ is a location-scale family with Z being the standard normal, $N(0,1)$. Exponential is a scale family with $\mu = 0$, $\sigma = \theta$. We can make it a location-scale family if we set

$$f(x|\mu,\sigma) = \begin{cases} \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } x > \mu; \\ 0 & \text{otherwise.} \end{cases}$$

Bernoulli, binomial, and Poisson are not location-scale families.

Example. Let X have uniform distribution over (θ_1, θ_2) so that

$$f(x|\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1} & \text{if } \theta_1 < x < \theta_2; \\ 0 & \text{otherwise.} \end{cases}$$

This is also a location-scale family, with a reparameterization.

Example. The Cauchy distribution specified by the density

$$f(x|\mu,\sigma) = \frac{1}{\pi}\frac{\sigma}{\sigma^2 + (x-\mu)^2}, \quad -\infty < x < \infty$$

is a location-scale family. It has several interesting properties. As $|x| \to \infty$, it tends to zero but at a much slower rate than the normal.

One can verify that $E(|X|^r) = \infty$ for $r = 1, 2, \ldots$ under any μ, σ. So Cauchy has no finite moment. However, Figure 1.1 shows remarkable similarity between the normal and Cauchy, except near the tails. The Cauchy density is much flatter at the tails than the normal, which means x's that deviate quite a bit from μ will appear in data from time to time. Such deviations from μ would be unusual under a normal model and so may be treated as outliers by a data analyst. It provides an important counter-example to the *law of large numbers* or *central limit theorem* when one has infinite moments. It also plays an important role in robustness studies.

Result. Any location-scale family of models is closed under location-scale transformations.

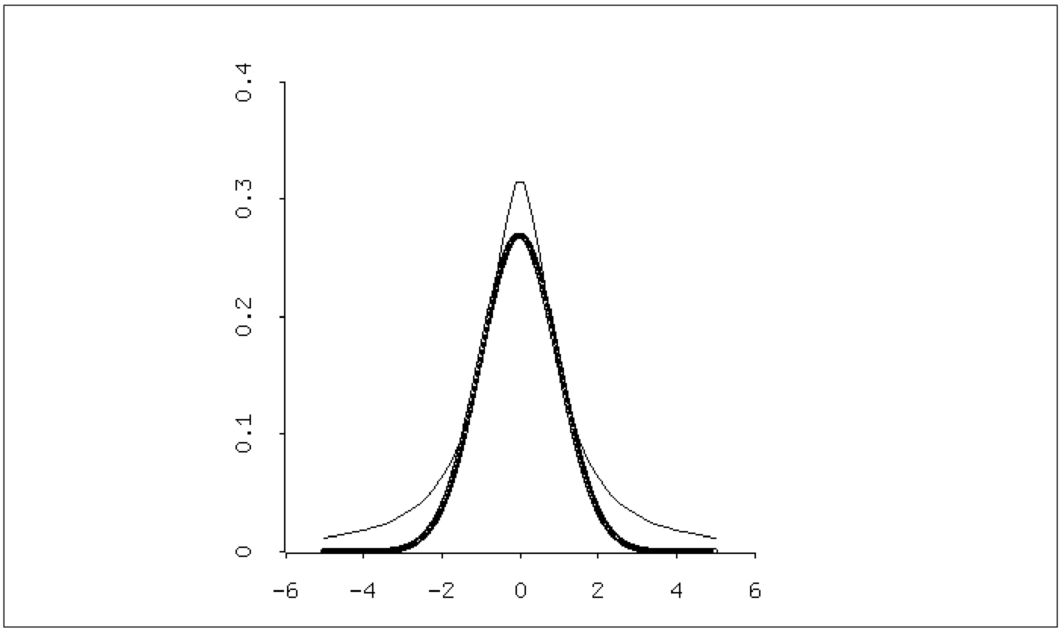

Figure 1: Densities of Cauchy(0, 1) and normal(0, 2.19).

Proof. Let $\mathcal{F}$ be a location-scale family. Then,

$$\mathcal{F} = \left\{ f(x|\mu,\sigma) = \frac{1}{\sigma} g\left(\frac{x-\mu}{\sigma}\right), -\infty < \mu < \infty, \sigma > 0 \right\},$$

where g is a density function. Let $Y = aX + b$. Take $a > 0$ for convenience. Then $X = (Y - b)/a$, so $dx = dy/a$, and hence

$$f_Y(y) = \frac{1}{a\sigma} g\left(\frac{\frac{y-b}{a} - \mu}{\sigma}\right) = \frac{1}{a\sigma} g\left(\frac{y - (a\mu + b)}{a\sigma}\right).$$

Therefore, $f_Y \in \mathcal{F}$.

Location-scale family is a special case of a Group family, a family of models which is closed under a group of transformations.

In our discussion we will confine ourselves to *parametric models*, in which case the parameter space is a nice subset of $\mathcal{R}^k$ for some k. Nonparametric models deal with much larger classes of models, such as all symmetric distributions or all symmetric unimodal distributions. Nonparametric methods (such as histograms) are quite different from what we discuss here. Even among parametric models, we restrict ourselves to models which satisfy the regularity conditions.