

Definition. Two distribution functions F, G have the same *type* if

$$G(x) = F(a + bx)$$

for some a, b . Then if $Y := (X - a)/b$, and $X \sim F$, then $Y \sim G$.

Stable laws.

The possible limit laws obtainable from centred and scaled random walks are the *stable* laws, which form the subclass S of SD . To within type (so we can take $a = 0$), they have two parameters, the *index* $\alpha \in (0, 2]$ and the *skewness parameter* $\beta \in [-1, 1]$. If $\alpha = 2$, the law is (standard) *normal*, and then $\sigma = 1, \nu = 0$. If $0 < \alpha < 2$, then $\sigma = 0$ and, for some $p \in [0, 1]$ and $q := 1 - p$ (the usual notation for Bernoulli trials $B(p)$), the Lévy measure has the form

$$d\nu = p \, dx/x^{1+\alpha} \quad \text{on } (0, \infty), \quad q \, dx/|x|^{1+\alpha} \quad \text{on } (-\infty, 0),$$

while the skewness parameter ('tail-balance parameter') is

$$\beta = p - q \quad (= 2p - 1)$$

(here $p + q = 1$, but this is a restriction of type, for convenience only – any value $p + q \in (0, \infty)$ will do). The CFs are

$$\phi(t) = \exp\left\{-\frac{1}{2}t^2\right\} \quad (\text{normal case, } \alpha = 2);$$

$$\phi(t) = \exp\left\{-|t|^\alpha(1 - i\beta(\operatorname{sgn} t) \tan \frac{1}{2}\pi\alpha)\right\} \quad (0 < \alpha < 1 \text{ or } 1 < \alpha < 2, -1 \leq \beta \leq 1);$$

$$\phi(t) = \exp\left\{-|t|(1 + i\beta(\operatorname{sgn} t) \frac{2}{\pi} \log |t|)\right\} \quad (\alpha = 1, -1 \leq \beta \leq 1).$$

If $\beta = 0$, the law is *symmetric* (X and $-X$ have the same distribution), and we obtain the *symmetric stable* laws with CFs

$$\phi(t) = \exp\{-|t|^\alpha\} \quad (0 < \alpha \leq 2).$$

Densities.

For $\alpha = 2$, we obtain the standard normal law, whose density $e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$

we know.

For $\alpha = 1$, we obtain the *symmetric Cauchy* law, whose density

$$f(x) = \frac{1}{\pi(1+x^2)}$$

we studied above.

For $\beta = +1$, $\alpha = \frac{1}{2}$, we obtain the *Lévy density* (Problems),

For other parameter values, there is no explicit formula, but one can obtain series expansions.

One-sided stable laws.

The Lévy measure is also called the *spectral measure*. When $\beta = +1$, $p = 1, q = 0$, all its mass is on the *positive* half-line – the *spectrally positive* case: only *positive jumps*, and decrease takes place continuously, rather than by jumping. Similarly for the *spectrally negative* case $\beta = -1$, $p = 0, q = 1$: all mass on the *negative* half-line; only negative jumps.

Stable laws and tails

It is a general property of id laws F that their tail behaviour is similar to that of their Lévy measures ν . Stable laws have finite a th moments for $a < \alpha$ and infinite a th moment for $a > \alpha$. Thus for $\alpha = 2$ (normal case) all moments are finite and the CF is entire; for $1 < \alpha < 2$ the mean exists but the variance does not; for $0 < \alpha < 1$ the mean does not exist. Such behaviour is described as having *heavy tails*. These are important, in at least two areas:

1. *Insurance*. It is the *large claims* that are dangerous for an insurance company – indeed, potentially lethal. The frequency of large claims is governed by the tail decay.
2. *Finance*. The standard benchmark model of mathematical finance, the *Black-Scholes(-Merton) model* has normal (actually, log-normal) tails. But most real financial data show much fatter tail behaviour than this. Stable laws have been used to model tails of financial data. So too have Student t -distributions (which, unlike stable laws, are not restricted to $\alpha \leq 2$).

IV. NORMAL DISTRIBUTION THEORY

1. Regression

In regression (see e.g. [BF]), we have data y_1, \dots, y_n , arranged as a column n -vector y . We seek to explain the data parsimoniously, in terms of p parameters β_1, \dots, β_p (arranged as a column p -vector β), via linear combinations of explanatory variables (predictor variables, covariates, regressors, ...), plus some error, which we model as an n -vector ϵ , whose components ϵ_i are assumed iid $N(0, \sigma^2)$. Then the *model equation* is

$$y = A\beta + \epsilon. \quad (ME)$$

Here the matrix A is $n \times p$; $p \ll n$ ("p is much less than n") – n is the sample size (the larger the better), p the number of parameters (as small as possible, by the Principle of Parsimony); A is called the *design matrix*. We restrict attention to the case when A has *full rank*, p (otherwise, eliminate superfluous regressors to reduce to this). From the model equation

$$y_i = \sum_{j=1}^p a_{ij}\beta_j, \quad \epsilon_i \text{ iid } N(0, \sigma^2),$$

the likelihood is

$$\begin{aligned} L &= \frac{1}{\sigma^n 2\pi^{\frac{1}{2}n}} \cdot \prod_{i=1}^n \exp\left\{-\frac{1}{2}(y_i - \sum_{j=1}^p a_{ij}\beta_j)^2 / \sigma^2\right\} \\ &= \frac{1}{\sigma^n 2\pi^{\frac{1}{2}n}} \cdot \exp\left\{-\frac{1}{2} \sum_{i=1}^n (y_i - \sum_{j=1}^p a_{ij}\beta_j)^2 / \sigma^2\right\}, \end{aligned}$$

and the log-likelihood is

$$\ell := \log L = \text{const} - n \log \sigma - \frac{1}{2} \left[\sum_{i=1}^n (y_i - \sum_{j=1}^p a_{ij}\beta_j)^2 \right] / \sigma^2.$$

As before, we use Fisher's Method of Maximum Likelihood, and maximise with respect to β_r : $\partial \ell / \partial \beta_r = 0$ gives

$$\sum_{i=1}^n a_{ir}(y_i - \sum_{j=1}^p a_{ij}\beta_j) = 0 \quad (r = 1, \dots, p),$$

or

$$\sum_{j=1}^p \left(\sum_{i=1}^n a_{ir}a_{ij} \right) \beta_j = \sum_{i=1}^n a_{ir}y_i.$$

Write $C = (c_{ij})$ for the $p \times p$ matrix

$$C := A^T A,$$

(called the *information matrix*), which we note is *symmetric*: $C^T = C$. Then

$$c_{ij} = \sum_{k=1}^n (A^T)_{ik} A_{kj} = \sum_{k=1}^n a_{ki} a_{kj}.$$

So this says

$$\sum_{j=1}^p c_{rj} \beta_j = \sum_{i=1}^n a_{ir} y_i = \sum_{i=1}^n (A^T)_{ri} y_i.$$

In matrix notation, this is

$$(C\beta)_r = (A^T y)_r \quad (r = 1, \dots, p),$$

or combining,

$$C\beta = A^T y, \quad C := A^T A. \quad (NE)$$

These are the *normal equations*.

As A has full rank, C is *positive definite* ($x^T C x > 0$ for all vectors $x \neq 0$) ([BF], Lemma 3.3), so we can solve the normal equations to obtain our *least-squares estimates* of β , namely

$$\hat{\beta} = C^{-1} A^T y.$$

Write

$$P := AC^{-1}A^T$$

for the *projection matrix* of A . Note that

$$P^2 = AC^{-1}A^T AC^{-1}A^T = AC^{-1}CC^{-1}A^T = AC^{-1}A^T = P,$$

so P is *idempotent*, i.e. is a *projection* (see e.g. [BF], Lemma 3.18). Also, as C is symmetric,

$$P^T = A(C^{-1})^T A^T = A(C^T)^{-1} A = AC^{-1}A^T = P:$$

P is *symmetric*, so is a *symmetric projection*. Similarly, so is $I - P$.

Call a linear transformation $P : V \rightarrow V$ a *projection onto V_1 along V_2* if V is the direct sum $V = V_1 \oplus V_2$, and if $x = (x_1, x_2)^T$ with $Px = x_1$. Then (check) $I - P$ is a projection onto V_2 along V_1 . Also

$$P(I - P) = P - P^2 = P - P = 0:$$

$P, I - P$ are *orthogonal projections*.