Chapter 4 Multivariate Distributions

Applied Multivariate Statistical Analysis

Course Instructor: Shuen-Lin Jeng

Department of Statistics
National Cheng Kung University

March, 2, 2021

4.1 Distribution and Density Function

Let $X = (X_1, X_2, \ldots, X_p)^T$ be a random vector. The cumulative distribution function (cdf) of X is defined by $F(x) = P(X \leq x) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p)$.

For continuous X, a nonnegative probability density function (pdf) f exists, that

$$F(x) = \int_{-\infty}^{x} f(u)du. \tag{4.1}$$

Note that

$$\int_{-\infty}^{\infty} f(u)du = 1.$$

4.1 Distribution and Density Function

For discrete X, the values of this random variable are concentrated on a countable or finite set of points $\{c_j\}_{j \in J}$, the probability of events of the form $\{X \in D\}$ can then be computed as

$$P(X \in D) = \sum_{\{j : c_j \in D\}} P(X = c_j).$$

If we partition as X as $X = (X_1, X_2)^T$ with $X_1 \in \mathbb{R}^k$ and $X_2 \in \mathbb{R}^{p-k}$, then the function

$$F_{X1}(x_1) = P(X_1 \leq x_1) = F(x_{11}, \ldots, x_{1k}, \infty, \ldots, \infty) \tag{4.2}$$

is called the *marginal cdf*. F $= F(x)$ is called the joint cdf.

4.1 Distribution and Density Function

For continuous X,

$$f_{X1}(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2. \tag{4.3}$$

The conditional pdf of X_2 given $X_1 = x_1$ is given as

$$f(x_2 \mid x_1) = \frac{f(x_1, x_2)}{f_{X1}(x_1)} \tag{4.4}$$

Definition 4.1

X_1 and X_2 are independent iff $f(x) = f(x_1, x_2) = f_{X1}(x_1) f_{X2}(x_2)$.

⋇ Different joint pdf's may have the same marginal pdf's.

4.1 Distribution and Density Function

An elegant concept of connecting marginals with joint cdfs is given by *copulae*.

For simplicity of presentation, we concentrate on the $p = 2$ dimensional case. A two-dimensional copula is a function $C : [0,1]^2 \to [0,1]$ with the following properties:

- For every $u \in [0,1]$: $\mathrm{C}(0,\ u) = \mathrm{C}(u,\ 0) = 0$.
- For every $u \in [0,1]$: $\mathrm{C}(u,\ 1) = u$ and $\mathrm{C}(1,\ u) = u$.
- For every $(u_1, u_2), (v_1, v_2) \in [0,1] \times [0,1]$ with $u_1 \le v_1$ and $u_2 \le v_2$:

$$C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \ge 0.$$

4.1 Distribution and Density Function

Theorem 4.1 (Sklar's theorem) *Let F be a joint distribution function with marginal distribution functions F_{X1} and F_{X2}. Then a copula C exists with*

$$F(x_1, x_2) = C\left\{F_{X_1}(x_1), F_{X_2}(x_2)\right\} \tag{4.5}$$

for every x_1, $x_2 \in \mathbb{R}$. If F_{X1} and F_{X2} are continuous, then C is unique. On the other hand, if C is a copula and F_{X1} and F_{X2} are distribution functions, then the function F defined by (4.5) is a joint distribution function with marginals F_{X1} and F_{X2}.

4.2 Moments and Characteristic Functions

Conditional Expectations

The conditional expectations are

$$\mathbf{E}(X_2 \mid x_1) = \int x_2 f(x_2 \mid x_1) dx_2 \text{ and } \mathbf{E}(X_1 \mid x_2) = \int x_1 f(x_1 \mid x_2) dx_1 \tag{4.27}$$

$$\mathbf{Var}(X_2 \mid X_1 = x_1) = \mathbf{E}(X_2 X_2^T \mid X_1 = x_1) - \mathbf{E}(X_2 \mid X_1 = x_1)\mathbf{E}(X_2^T \mid X_1 = x_1).$$

Using the conditional covariance matrix, the conditional correlations may be defined as

$$\rho_{X_2 X_3 \mid X_1 = x_1} = \frac{\mathbf{Cov}(X_2, X_3 \mid X_1 = x_1)}{\sqrt{\mathbf{Var}(X_2 \mid X_1 = x_1)\mathbf{Var}(X_3 \mid X_1 = x_1)}}.$$

4.2 Moments and Characteristic Functions

Properties of Conditional Expectations

Since $\mathbf{E}(X_2 \mid X_1 = x_1)$ is a function of x_1, say $h(x_1)$, we can define the random variable $h(X_1) = \mathbf{E}(X_2 \mid X_1)$. The same can be done when defining the random variable $\mathbf{Var}(X_2 \mid X_1)$.

$$\mathbf{E}(X_2) = \mathbf{E}\left\{\mathbf{E}(X_2 \mid X_1)\right\} \tag{4.28}$$

$$\mathbf{Var}(X_2) = \mathbf{E}\left\{\mathbf{Var}(X_2 \mid X_1)\right\} + \mathbf{Var}\left\{\mathbf{E}(X_2 \mid X_1)\right\} \tag{4.29}$$

4.2 Moments and Characteristic Functions

Theorem 4.3 *Let* $X_1 \in \mathbb{R}^k$ *and* $X_2 \in \mathbb{R}^{p-k}$ *and* $U = X_2 - \mathbf{E}(X_2 \mid X_1)$*. Then we have:*

1. $\mathbf{E}(U) = 0$
2. $\mathbf{E}(X_2 \mid X_1)$ *is the best approximation of* X_2 *by a function* $h(X_1)$ *of* X_1 *where h:* $\mathbb{R}^k \to \mathbb{R}^{p-k}$*. "Best" is the minimum mean squared error (MSE) sense, where*

$$MSE(h) = \mathbf{E}\left[\{X_2 - h(X_1)\}^T \{X_2 - h(X_1)\}\right].$$

4.2 Moments and Characteristic Functions

Characteristic Functions

The characteristic function (cf) of a random vector $X \in \mathbb{R}^p$ (respectively its density $f(x)$) is defined as

$$\varphi_X(t) = \mathbf{E}(e^{\mathbf{i}t^T X}) = \int e^{\mathbf{i}t^T X} dx, t \in \mathbb{R}^p,$$

where $\mathbf{i}$ is the complex unit: $\mathbf{i}^2 =$ -1.

$$\varphi_X(0) = 1\text{and } |\varphi_X(t)| \leq 1. \tag{4.30}$$

If φ is absolutely integrable, i.e., the integral $\int_{-\infty}^{\infty} |\varphi(x)| \, dx$ exists and is finite, then,

$$f(x) = \frac{1}{(2\pi)^p} \int_{-\infty}^{\infty} e^{-\mathbf{i}t^T x} \varphi_X(t) dt \tag{4.31}$$

4.2 Moments and Characteristic Functions

The characteristic function can recover all the cross-product moments of any order: $\forall j_k \geq 0, k = 1, \ldots, p$ and for $t = (t_1 \cdot \ldots \cdot t_p)^T$ we have

$$\mathbf{E}(X_1^{j_1} \cdot \ldots \cdot X_p^{jp}) = \frac{1}{\mathbf{i}^{j1+\ldots+jp}} \left[\frac{\partial \varphi_X(t)}{\partial t_1^{j1} \ldots \partial t_p^{jp}} \right]_{t=0}. \tag{4.35}$$

Table 4.1 Characteristic functions for some common distributions

	pdf	cf
Uniform	$f(x) = \mathrm{I}(x \in [a, b])/(b-a)$	$\varphi_X(t) = (e^{\mathbf{i}bt} - e^{\mathbf{i}at})/(b-a)\mathbf{i}t$
$N_1(\mu, \sigma^2)$	$f(x) = (2\pi\sigma^2)^{-1/2}\exp\{-(x-\mu)^2/2\sigma^2\}$	$\varphi_X(t) = e^{\mathbf{i}\mu t - \sigma^2 t^2/2}$
$\chi^2(n)$	$f(x) = \mathrm{I}(x > 0)x^{n/2-1}e^{-x/2}/\{\Gamma(n/2)2^{n/2}\}$	$\varphi_X(t) = (1 - 2\mathbf{i}t)^{-n/2}$
$N_p(\mu, \Sigma)$	$f(x) = \lvert 2\pi\Sigma\rvert^{-1/2}\exp\{-(x-\mu)^\top \Sigma(x-\mu)/2\}$	$\varphi_X(t) = e^{\mathbf{i}t^\top\mu - t^\top\Sigma t/2}$

4.2 Moments and Characteristic Functions

Theorem 4.4 (Cramer-Wold) *The distribution of $X \in \mathbb{R}^p$ is completely determined by the set of all(one-dimensional) distributions of $t^T X$ where $t \in \mathbb{R}^p$.*

Cumulant functions

Moments $m_k = \int x^k f(x)dx$ often help in describing distributional characteristics.

$\mu = m_1$ and $\sigma^2 = m_2 - m_1^2$. Skewness γ_3 and kurtosis γ_4 are defined as

$$\gamma_3 = \mathbf{E}(X-\mu)^3/\sigma^3$$

$$\gamma_4 = \mathbf{E}(X-\mu)^4/\sigma^4 \tag{4.40}$$

4.3 Transformations

Asking for the pdf of Y when

$$X = u(Y) \tag{4.43}$$

for a one-to-one transformation u: $\mathbb{R}^p \to \mathbb{R}^p$. Define the Jacobian of u as

$$\mathcal{J} = \left(\frac{\partial x_i}{\partial y_j}\right) = \left(\frac{\partial u_i(y)}{\partial y_j}\right).$$

and let abs $(|\mathcal{J}|)$ be the absolute value of the determinant of this Jacobian. The pdf of Y is given by

$$f_Y(y) = abs\left(|\mathcal{J}|\right) \cdot f_X\left\{u(y)\right\} \tag{4.44}$$

4.3 Transformations

This introductory example is a special case of

$$Y = \mathcal{A}X + b \text{ , where } \mathcal{A} \text{ is nonsingular.}$$

The inverse transformation is

$$X = \mathcal{A}^{-1}(Y - b)$$

Therefore

$$\mathcal{J} = \mathcal{A}^{-1}$$

and hence

$$f_Y(y) = abs(|\mathcal{A}|^{-1}) f_x \left\{ \mathcal{A}^{-1}(y - b) \right\}. \tag{4.45}$$

4.4 The Multinormal Distribution

The multinormal distribution with mean μ and covariance $\sum > 0$ has the density

$$f(x) = \left|2\pi \sum\right|^{-1/2} exp\left\{-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)\right\}. \qquad (4.47)$$

We write $\mathrm{X} \sim N_p(\mu, \sum)$.

Theorem 4.5 *Let $X \sim N_p(\mu, \sum)$ and $Y = \sum^{-1/2}(X-\mu)$ (Mahalanobis transformation). Then*

$$Y \sim N_p(0, \mathcal{I}_p),$$

i.e., the elements $Y_j \in \mathbb{R}$ are independent, one-dimensional $N(0,1)$ variables.

4.4 The Multinormal Distribution

Theorem4.6 *Let* $X \sim N_p(\mu, \sum)$ *and* $\mathcal{A}(p \times p)$, $c \in \mathbb{R}^p$, *where* $\mathcal{A}$ *is nonsingular. Then* $Y = \mathcal{A}X + c$ *is again a p-variate Normal, i.e.,*

$$Y \sim N_p\left(\mathcal{A}\mu + c, \mathcal{A}\sum\mathcal{A}^T\right). \tag{4.50}$$

Geometry of the $N_p(\mu, \sum)$ Distribution

Theorem 4.7 *If* $X \sim N_p(\mu, \sum)$,
then the variable $U = (X - \mu)^T \sum^{-1}(X - \mu)$ *has a* χ^2_p *distribution.*

Theorem 4.8 *The characteristic function (cf) of a multinormal* $N_p(\mu, \sum)$ *is given by*

$$\varphi_X(t) = exp\left(\mathbf{i}t^T\mu - \frac{1}{2}t^T\sum t\right). \tag{4.52}$$

4.4 The Multinormal Distribution

Singular Normal Distribution

Suppose that we have rank$(\sum) = k < p$,where p is the dimension of X. We define the (singular) density of X with the aid of the G-Inverse $\sum^{-}$ of $\sum$,

$$f(x) = \frac{(2\pi)^{-k/2}}{(\lambda_1 \dots \lambda_k)^{1/2}} exp \left\{ -\frac{1}{2}(x-\mu)^T \sum^{-} (x-\mu) \right\} \tag{4.53}$$

where

1. x lies on the hyperplane $\mathcal{N}^T(x-\mu) = 0$ with $\mathcal{N}(p \times (p-k))$: $\mathcal{N}^T \sum = 0$ and $\mathcal{N}^T \mathcal{N} = \mathcal{I}_k$.
2. $\sum^{-}$ is the G-Inverse of $\sum$, and $\lambda_1, \dots, \lambda_k$ are the nonzero eigenvalues of $\sum$.

4.4 The Multinormal Distribution

What is the connection to a multinormal with k-dimensions? If

$$Y \sim N_k(0, \Lambda_1) \text{ and } \Lambda_1 = diag(\lambda_1, \ldots, \lambda_k), \tag{4.54}$$

then an orthogonal matrix $\mathcal{B}(p \times k)$ with $\mathcal{B}^{\mathcal{T}}\mathcal{B} = \mathcal{I}_k$ exists that means $X = \mathcal{B}Y + \mu$ where X has a singular pdf of the form (4.53).

4.4 The Multinormal Distribution

Gaussian Copula

Gaussian or *normal copula*,

$$C_\rho(u, v) = \int_{-\infty}^{\Phi_1^{-1}(u)} \int_{-\infty}^{\Phi_2^{-1}(v)} f_\rho(x_1, x_2) dx_2 dx_1, \tag{4.55}$$

see Embrechts et al.(1999). In(4.55), f_ρ denotes the bivariate normal density function with correlation ρ for $n = 2$. The functions Φ_1 and Φ_2 in (4.55) refer to the corresponding one-dimensional standard normal cdfs of the marginals. In the case of vanishing correlation, $\rho = 0$, the Gaussian copula becomes

$$C_0(u, v) = \int_{-\infty}^{\Phi_1^{-1}(u)} f_{X1}(x_1) dx_1 \int_{-\infty}^{\Phi_2^{-1}(v)} f_{X2}(x_2) dx_2 = uv = \prod(u, v).$$

4.5 Sampling Distributions and Limit Theorems

Theorem 4.9 *Let $X_1, \ldots, X_n$ be i.i.d. with $X_i \sim N_p(\mu, \sum)$. Then $\bar{x} \sim N_p\left(\mu, n^{-1} \sum\right)$.*

Theorem 4.10 (Central Limit Theorem (CLT)) *Let $X_1, X_2, \ldots, X_n$ be i.i.d. with $X_i \sim (\mu, \sum)$. Then the distribution of $\sqrt{n}(\bar{x} - \mu)$ is asymptotically $N_p(0, \sum)$, i.e.,*

$$\sqrt{n}(\bar{x} - \mu) \xrightarrow{\mathcal{L}} N_p(0, \sum) \text{ as } n \to \infty$$

Corollary 4.1 *If $\hat{\sum}$ is a consistent estimate for $\sum$, then the CLT still holds, namely,*

$$\sqrt{n}\hat{\sum}^{-1/2}(\bar{x} - \mu) \xrightarrow{\mathcal{L}} N_p(0, \mathcal{I}) \text{ as } n \to \infty.$$

4.5 Sampling Distributions and Limit Theorems

Transformation of Statistics

Theorem 4.11 *If* $\sqrt{n}(t-\mu) \xrightarrow{\mathcal{L}} N_p(0, \sum)$ *and if* $f = (f_1, \ldots, f_q)^T : \mathbb{R}^p \to \mathbb{R}^q$ *are real-valued functions which are differentiable at* $\mu \in \mathbb{R}^p$*, then* $f(t)$ *is asymptotically normal with mean* $f(\mu)$ *and covariance* $\mathcal{D}^T \sum \mathcal{D}$*, i.e.,*

$$\sqrt{n}\left\{f(t) - f(\mu) \xrightarrow{\mathcal{L}} N_q(0, \mathcal{D}^T \sum \mathcal{D}) \text{ for } n \to \infty\right\} \tag{4.56}$$

where

$$\mathcal{D} = \left(\frac{\partial f_j}{\partial t_i}\right)(t)\bigg|_{t=\mu}$$

is the $(p \times q)$ matrix of all partial derivatives.

4.5 Sampling Distributions and Limit Theorems

Example 4.20 We are interested in seeing how $f(\bar{x}) = \bar{x}^T \mathcal{A}\bar{x}$ behaves asymptotically respect to the quadratic cost function of μ. $f(\mu) = \mu^T \mathcal{A}\mu$, where $\mathcal{A} > 0$.

$$D = \frac{\partial f(\bar{x})}{\partial \bar{x}}\bigg|_{\bar{x}=\mu} = 2\mathcal{A}\mu.$$

By Theorem 4.11 , we have

$$\sqrt{n}(\bar{x}^T \mathcal{A}\bar{x} - \mu^T \mathcal{A}\mu) \xrightarrow{\mathcal{L}} N_1(0, 4\mu^T \mathcal{A} \sum \mathcal{A}\mu).$$

4.6 Heavy-Tailed Distributions

A distribution is called heavy-tailed if it has higher probability density in its tail area compared with a normal distribution with same mean μ and variance σ^2.

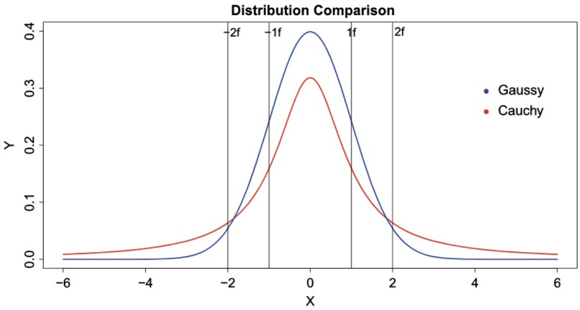

Fig. 4.6 Comparison of the pdf of a standard Gaussian (blue) and a Cauchy distribution (red) with location parameter 0 and scale parameter 1 MVAgausscauchy

4.6 Heavy-Tailed Distributions

Generalized Hyperbolic Distribution

The generalized hyperbolic distribution was introduced by Barndorff-Nielsen and at first applied to model grain size distributions of windblown sands. Today one of its most important uses is in stock price modeling and market risk measurement. The name of the distribution is derived from the fact that its log-density forms a hyperbola, while the log-density of the normal distribution is a parabola (Fig. 4.7).

4.6 Heavy-Tailed Distributions

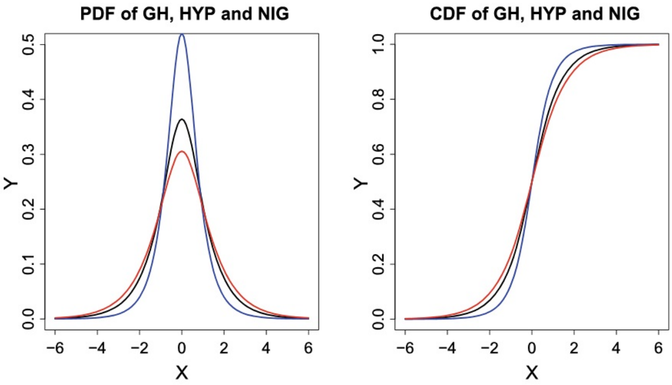

Fig. 4.7 pdf (left) and cdf (right) of GH($\lambda = 0.5$, black), HYP(red), and NIG (blue) with $\alpha = 1, \beta = 0, \delta = 1, \mu = 0$ MVAghdis

4.6 Heavy-Tailed Distributions

The density of a one-dimensional generalized hyperbolic (GH) distribution for $x \in \mathbb{R}$ is

$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu)$

$$= \frac{(\sqrt{\alpha^2 - \beta^2}/\delta)^\lambda}{\sqrt{2\pi} K_\lambda(\delta\sqrt{\alpha^2 - \beta^2})} \frac{K_{\lambda-1/2}\left\{\alpha\sqrt{\delta^2 + (x-\mu)^2}\right\}}{(\sqrt{(\delta^2 + (x-\mu)^2)}/\alpha)^{1/2-\lambda}} e^{\beta(x-\mu)} \tag{4.57}$$

where K_λ is a modified Bessel function of the third kind with index λ

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy \tag{4.58}$$

4.6 Heavy-Tailed Distributions

The domain of variation of the parameters is $\mu \in \mathbb{R}$ and

$$\delta \geq 0, |\beta| < \alpha, \text{ if } \lambda > 0$$

$$\delta > 0, |\beta| < \alpha, \text{ if } \lambda = 0$$

$$\delta > 0, |\beta| \leq \alpha, \text{ if } \lambda < 0$$

For $\lambda = 1$, we obtain the hyperbolic distributions (HYP)

For $\lambda = -1/2$, we obtain the normal-inverse Gaussian distribution(NIG).

4.6 Heavy-Tailed Distributions

Student's t-distribution

Let X be a normally distributed random variable with mean μ and variance σ^2, and Y be the random variable such that Y^2/σ^2 has a chi-square distribution with n degrees of freedom. Assume that X and Y are independent, then

$$t \overset{def}{=} \frac{X\sqrt{n}}{Y} \tag{4.63}$$

is distributed as Student's t with n degrees of freedom.

4.6 Heavy-Tailed Distributions

The t-distribution has the following density function:

$$f_t(x; n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \tag{4.64}$$

where n is the number of degrees of freedom, $-\infty < x < \infty$, and Γ is the gamma function:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \tag{4.65}$$

4.6 Heavy-Tailed Distributions

The mean, variance, skewness, and kurtosis of Student's *t*-distribution (n $>$ 4) are

$$\mu = 0$$

$$\sigma^2 = \frac{n}{n-2}$$

$$\text{Skewness } = 0$$

$$\text{Kurtosis } = 3 + \frac{6}{n-4}$$

4.6 Heavy-Tailed Distributions

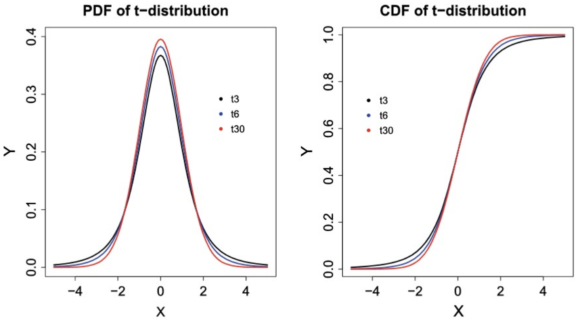

Fig. 4.8 pdf (left) and cdf (right) of t-distribution with different degrees of freedom (t3 stands for t-distribution with degree of freedom 3) MVAtdis

Student's t-distribution approaches the normal distribution as n increases.

4.6 Heavy-Tailed Distributions

Laplace distribution

The Laplace distribution can be defined as the distribution of differences between two independent variates with identical exponential distributions. Therefore, it is also called the double exponential distribution (Fig. 4.9). The Laplace distribution with mean μ and scale parameter θ has the pdf

$$f_{Laplace}(x;\mu,\theta) = \frac{1}{2\theta} e^{-\frac{|x-\mu|}{\theta}} \tag{4.67}$$

and the cdf

$$F_{Laplace}(x;\mu,\theta) = \frac{1}{2}\left\{1 + sign(x-\mu)\left(1 - e^{-\frac{|x-\mu|}{\theta}}\right)\right\} \tag{4.68}$$

4.6 Heavy-Tailed Distributions

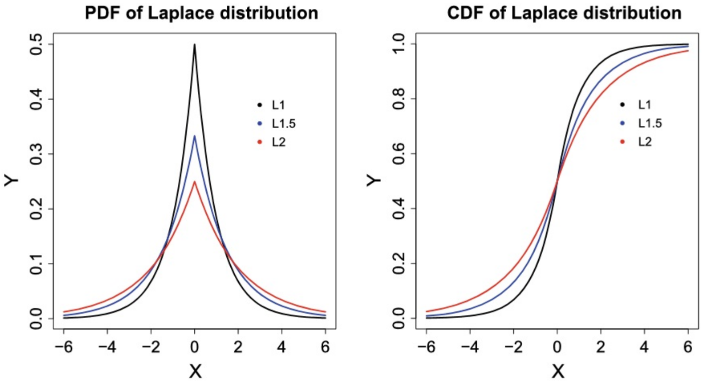

Fig. 4.9 pdf (left) and cdf (right) of Laplace distribution with zero mean and different scale parameters (L1 stands for Laplace distribution with $\theta = 1$) MVAlaplacedis

4.6 Heavy-Tailed Distributions

The mean, variance, skewness, and kurtosis of Laplace distribution are

$$\mu = \mu$$

$$\sigma^2 = 2\theta^2$$

$$\text{Skewness } = 0$$

$$\text{Kurtosis } = 6$$

4.6 Heavy-Tailed Distributions

Cauchy distribution

The general formula for the pdf and cdf of the Cauchy distribution is

$$f_{Cauchy}(x; m, s) = \frac{1}{s\pi} \frac{1}{1 + \left(\frac{x-m}{s}\right)^2} \tag{4.71}$$

$$F_{Cauchy}(x; m, s) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-m}{s}\right) \tag{4.72}$$

where m and s are location and scale parameter, respectively.

4.6 Heavy-Tailed Distributions

The case in the above example where $m = 0$ and $s = 1$ is called the standard Cauchy distribution with pdf and cdf as following:

$$f_{Cauchy}(x) = \frac{1}{\pi(1 + x^2)} \tag{4.73}$$

$$F_{Cauchy}(x; m, s) = \frac{1}{2} + \frac{\arctan(x)}{\pi} \tag{4.74}$$

The mean, variance, skewness, and kurtosis of Cauchy distribution are all undefined. But it has mode and median, both equal to the location parameter m. (Fig. 4.11).

4.6 Heavy-Tailed Distributions

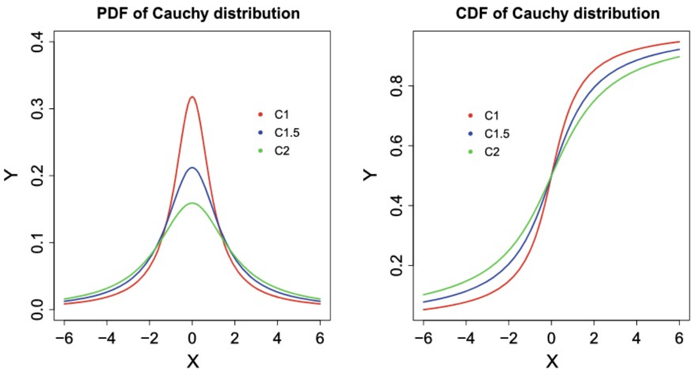

Fig. 4.11 pdf (left) and cdf (right) of Cauchy distribution with $m = 0$ and different scale parameters (C1 stands for Cauchy distribution with $s = 1$) MVAcauchy

4.6 Heavy-Tailed Distributions

Mixture Model

Table 4.2 Basic statistics of t, Laplace, and Cauchy distribution

	t	Laplace	Cauchy
Mean	0	μ	Not defined
Variance	$\frac{n}{n-2}$	$2\theta^2$	Not defined
Skewness	0	0	Not defined
Kurtosis	$3+\frac{6}{n-4}$	6	Not defined

4.6 Heavy-Tailed Distributions

Multivariate Generalized Hyperbolic Distribution

The multivariate Generalized Hyperbolic Distribution (GH_d) has the following pdf:

$f_{GH_d}(x; \lambda, \alpha, \beta, \delta, \Delta, \mu) =$

$$a_d \frac{K_{\lambda - \frac{d}{2}} \left\{ \alpha \sqrt{\delta^2 + (x-\mu)^T \Delta^{-1} (x-\mu)} \right\}}{\left\{ \alpha^{-1} \sqrt{\delta^2 + (x-\mu)^T \Delta^{-1} (x-\mu)} \right\}^{\frac{d}{2} - \lambda}} e^{\beta^T (x-\mu)} \tag{4.85}$$

$a_d = a_d(\lambda, \alpha, \beta, \delta, \Delta) =$

$$\frac{\left(\sqrt{\alpha^2 - \beta^T \Delta \beta} / \delta \right)^{\lambda}}{(2\pi)^{\frac{d}{2}} K_\lambda \delta \sqrt{\alpha^2 - \beta^T \Delta \beta}}, \tag{4.86}$$

4.6 Heavy-Tailed Distributions

Multivariate t-distribution

If X and Y are independent and distributed as $N_p(\mu, \sum)$ and χ^2_n, respectively, and $X\sqrt{n/Y} =$ t - μ, then the pdf of t is given by

$f_t(t; n, \sum, \mu) =$

$$\frac{\Gamma\left\{(n+p)/2\right\}}{\Gamma(n/2)n^{p/2}\pi^{p/2}|\sum|^{1/2}\left\{1+\frac{1}{n}(t-\mu)^T\sum^{-1}(t-\mu)\right\}^{(n+p)/2}} \tag{4.96}$$

The distribution of t is the noncentral t-distribution with n degrees of freedom and the noncentrality parameter μ, Giri(1996).

4.6 Heavy-Tailed Distributions

Multivariate Mixture Model

A multivariate mixture model comprises multivariate distributions, e.g., the pdf of a multivariate Gaussian distribution can be written as

$$f(x) = \sum_{l=1}^{L} \frac{w_l}{|2\pi \sum_l|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_l)^T \sum^{-1}(x-\mu_l)} \tag{4.103}$$

4.6 Heavy-Tailed Distributions

Generalized Hyperbolic Distribution The GH distribution has an exponential decaying speed

$$f_{GH}(x; \lambda, \alpha, \beta, \delta, \mu = 0) \sim x^{\lambda-1} e^{-(\alpha-\beta)x} \text{ as } x \to \infty, \tag{4.104}$$

Furthermore, we used one important subclass of the GH distribution: the normal-inverse Gaussian(NIG) distribution with $\lambda = -\frac{1}{2}$ introduced above.

4.6 Heavy-Tailed Distributions

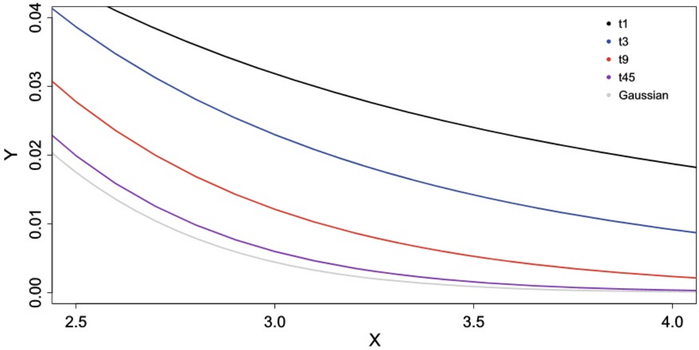

Fig. 4.13 Tail comparison of t-distributions (pdf) MVAtdistail

4.6 Heavy-Tailed Distributions

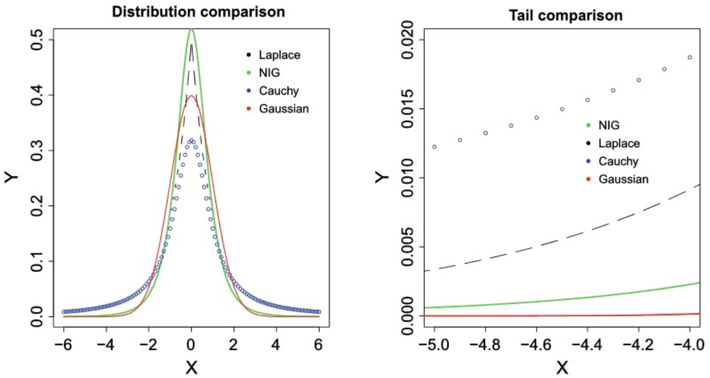

Fig. 4.15 Graphical comparison of the NIG, Laplace, Cauchy, and standard normal distribution

MVAghadatail