Cauchy Distribution

\[ f(x) = \frac{1}{\pi} \left[ \frac{\gamma}{(x - x_0)^2 + \gamma^2} \right] \]

where \( x_0 \) is the location parameter, \( \gamma \) is the scale parameter. The standard Cauchy distribution is the case where \( x_0 = 0 \) and \( \gamma = 1 \).

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

Note that the \( t \) distribution with \( v = 1 \) becomes a standard Cauchy distribution. Also note that the mean and variance of the Cauchy distribution don’t exist.

### 2.3 Representations of A Probability Distribution

**Survival Function**

\[ S(x) = 1 - F(x) = \text{Prob}[X \geq x] \]

where \( X \) is a continuous random variable.

**Hazard Function (Failure rate)**

\[ h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)} \]

Let \( f(t) = \lambda e^{-\lambda t} \) (exponential density function), then we have

\[ h(t) = \frac{f(t)}{S(t)} = \lambda \]
which implies that the hazard rate is a constant with respect to time. However for Weibull distribution or log normal distribution, the hazard function is not a constant any more.

**Moment Generating Function (mgf)**  The mgf of a random variable $X$ is

$$M_x(t) = E(e^{tx})$$

for $t \in \mathbb{R}$

Note that mgf is an alternate definition of probability distribution. Hence there is one for one relationship between the pdf and mgf. However mgf does not exist sometimes. For example, the mgf for the Cauchy distribution is not able to be defined.

**Characteristic Function (cf)**  Alternatively, the following characteristic function is used frequently in Finance to define probability function. Even when the mgf does not exist, cf always exist.

$$\phi(t) = E(e^{itx})$$

For example, the cf for the Cauchy distribution is $\exp(x_0 it - \gamma |t|)$.

**Cumulants**  The cumulants $\kappa_n$ of a random variable $X$ are defined by the cumulant generating function which is the logarithm of the mgf.

$$g(t) = \log [E(e^{tx})]$$

Then, the cumulants are given by

$$\kappa_1 = \mu = g'(0)$$
$$\kappa_2 = \sigma^2 = g''(0)$$
$$\kappa_n = g^{(n)}(0)$$
2.4 Joint Distributions

The joint distribution for $\mathbf{x}$ and $\mathbf{y}$ denoted $\mathbf{\varphi}(\mathbf{x}, \mathbf{y})$ is defined as

$$\text{Prob}(a \leq x \leq b, c \leq y \leq d) = \left\{ \begin{array}{ll}
\sum_{a}^{b} \sum_{c}^{d} f(x, y) \\
\int_{a}^{b} \int_{c}^{d} f(x, y) \, dx \, dy
\end{array} \right.$$ 

Consider the following bivariate normal distribution as an example.

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp \left( -\frac{1}{2} \left[ \frac{x - \mu_x}{\sigma_x} \cdot \frac{y - \mu_y}{\sigma_y} - 2\rho \frac{x - \mu_x}{\sigma_x} \frac{y - \mu_y}{\sigma_y} \right] / (1 - \rho^2) \right)$$

where

$$\varepsilon_x = \frac{x - \mu_x}{\sigma_x}, \quad \varepsilon_y = \frac{y - \mu_y}{\sigma_y}, \quad \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

Suppose that $\sigma_x = \sigma_y = 1, \mu_x = \mu_y = 0, \sigma_{xy} = 0.5$. Then we have

$$f(x, y) = \frac{1}{2\pi\sqrt{1 - 0.5^2}} \exp \left( -\frac{1}{2} \left[ x^2 + y^2 - xy \right] / (1 - 0.5^2) \right)$$

We denote

$$\left( \begin{array}{c}
x \\
y
\end{array} \right) \sim N \left( \left[ \begin{array}{c}
\mu_x \\
\mu_y
\end{array} \right], \left[ \begin{array}{cc}
\sigma_x^2 & \sigma_{xy} \\
\sigma_{xy} & \sigma_y^2
\end{array} \right] \right)$$
Marginal distribution  It is defined as

\[ \text{Prob}(x = x_0) = \sum_y \text{Prob}(x = x_0|y = y_0) \text{Prob}(y = y_0) = f_x(x) = \left\{ \begin{array}{l}
\sum_y f(x, y) \\
\int_y f(x, s) ds
\end{array} \right. \]

Note that

\[ f(x, y) = f_x(x) f_y(y) \text{ iff } x \text{ and } y \text{ are independent} \]

Also note that if \( x \) and \( y \) are independent, then

\[ F(x, y) = F_x(x) F_y(y) \]

alternatively

\[ \text{Prob}(x \leq x_0, y \leq y_0) = \text{Prob}(x \leq x_0) \text{Prob}(y \leq y_0) \]

For a bivariate normal distribution case, the marginal distribution is given by

\[ f_x(x) = N(\mu_x, \sigma_x^2) \]
\[ f_y(y) = N(\mu_y, \sigma_y^2) \]

Expectations in a joint distribution  Mean:

\[ E(x) = \left\{ \begin{array}{l}
\sum_x x f_x(x) = \sum_x x \sum_y f(x, y) \\
\int_x x f_x(x) dx = \int_x \int_y x f(x, y) dydx
\end{array} \right. \]

Variance: See B-50.

Covariance and Correlation

\[ \text{Cov}[x, y] = E[(x - \mu_x)(y - \mu_y)] = \sigma_{xy} \]
\[ \text{Corr}(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \]

\[ V(x + y) = V(x) + V(y) + 2 \text{Cov}(x, y) \]
2.5 Conditioning in a bivariate distribution

\[ f(y|x) = \frac{f(x,y)}{f_x(x)} \]

For a bivariate normal distribution case, the conditional distribution is given by

\[ f(y|x) = N(\alpha + \beta x, \sigma_y^2 (1 - \rho^2)) \]

where \( \alpha = \mu_y - \beta \mu_x, \ \beta = \sigma_{xy}/\sigma_x^2 \).

If \( \rho = 0 \), then \( y \) and \( x \) are independent.

**Regression: The Conditional Mean**  The conditional mean is the mean of the conditional distribution which is defined as

\[ E(y|x) = \left\{ \begin{array}{l}
\sum_y y f(y|x) \\
\int_y y f(y|x) \, dy
\end{array} \right. \]

The conditional mean function \( E(y|x) \) is called the regression of \( y \) on \( x \).

\[ y = E(y|x) + (y - E(y|x)) \]

\[ = E(y|x) + \varepsilon \]

Example:

\[ y = a + bx + \varepsilon \]

Then

\[ E(y|x) = a + bx. \]

**Conditional Variance**

\[ V(y|x) = E[(y - E(y|x))^2 |x] \]

\[ = E(y^2|x) - E(y|x)^2 \]
2.6 The Multivariate Normal Distribution

Let $x = (x_1, ..., x_n)'$ and have a multivariate normal distribution. Then we have

$$f(x) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu)\right)$$

1. If $x \sim N(\mu, \Sigma)$, then

   $$Ax + b \sim N(A\mu + b, A\Sigma A')$$

2. If $x \sim N(\mu, \Sigma)$, then

   $$(x - \mu)' \Sigma^{-1} (x - \mu) \sim \chi^2_n$$

3. If $x \sim N(\mu, \Sigma)$, then

   $$\Sigma^{-1/2} (x - \mu) \sim N(0, I_n)$$
2.7 Sample Questions

Q1: Write down the definitions of skewness and kurtosis. What is the value of skewness for the symmetric distribution.

Q2: Let \( x_i \sim N(0, \sigma^2) \) for \( i = 1, ..., n \). Further assume that \( x_i \) is independent each other. Then

1. \( x_i^2 \sim \)
2. \( \sum_{i=1}^{n} x_i^2 \sim \)
3. \( \sum_{i=1}^{n} \frac{x_i^2}{\sigma^2} \sim \)
4. \( \frac{x_i^2}{x_j^2} \sim \)
5. \( \frac{x_i^2 + x_j^2}{x_3^2} \sim \)
6. \( \frac{x_1}{x_2^2} \sim \)
7. \( \frac{x_1}{x_2^2 + x_3} \sim \)

Q3: Write down the standard normal density

Q4: Let \( x \sim LN(\mu, \sigma^2) \).

1. \( \ln(x) \sim \)
2. Prove that \( y = ax \sim LN(\ln a + \mu, \sigma^2) \).
3. Prove that \( y = 1/x \sim LN(-\mu, \sigma^2) \).

Q5: Write down the density function of the Gamma distribution

1. Write down the values of \( p \) and \( \lambda \) when Gamma = \( \chi^2 \)
2. Write down the values of \( p \) and \( \lambda \) when Gamma = exponential distribution
Q6: Write down the density function of the logistic distribution.

Q7: Write down the density function of Cauchy distribution. Write down the value of \( v \) when Cauchy=\( t \) distribution

Q8: Write down the definition of Moment Generating and Characteristic function.

Q9: Suppose that \( \mathbf{x} = (x_1, x_2, x_3)' \) and \( \mathbf{x} \sim N(\mu, \Sigma) \)

1. Write down the normal density in this case.

2. \( \mathbf{y} = \mathbf{Ax} + \mathbf{b} \sim \)

3. \( \mathbf{z} = \Sigma^{-1} (\mathbf{x} - \mathbf{c}) \sim \) where \( \mathbf{c} \neq \mu \).
3 Estimation and Inference

3.1 Definitions

1. Random variable and constant: A random variable is believed to change over time across individual. Constant is believed not to change either dimension. It becomes an issue in the panel data.

\[ x_{it} = a_i + x_{it}^o \]

Here we decompose \( x_{it} \) \((i = 1, ..., N; t = 1, ..., T)\) into its mean (time invariant) and time varying components. Now is \( a_i \) random or constant. According to the definition of random variables, \( a_i \) can be a constant since it does not change over time. However, if \( a_i \) has a pdf, then it becomes a random variable.

2. IID: independent, identically distributed: Consider the following sequence

\[ x = (x_1, x_2, x_2) = (1, 2, 3) \]

Now we are asking if each number is a random variable or constant. If they are random, then we have to ask the pdf of each number. Suppose that

\[ x_i \sim N \left(0, \sigma_i^2\right), \]

Now we have to know that \( x_i \) is an independent event. If they are independent, then next we have to know \( \sigma_i^2 \) is identical or not. Typical assumption is IID.

3. Mean:

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

4. Standard error (deviation)

\[ s_x = \left[ \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right]^{1/2} \]

5. Covariance

\[ s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \]
6. Correlation

\[ \gamma_{xy} = \frac{s_{xy}}{s_x s_y} \]

**Population values and estimates**

\[ y_i = bx_i + u_i \]

\[ u_i \sim iidN(\mu, \sigma^2) \]

1. Estimate: it is a statistic computed from a sample \( \hat{b} \). Sample mean is the statistic for the population mean \( \mu \).

2. Standard deviation and error: \( \sigma \) is the standard deviation and \( s_x \) is the standard error of the population.

3. Regression error and residual: \( u_i \) is the error, \( \hat{u}_i \) is the residual

4. Estimator: a rule or method for using the data to estimate the parameter. “OLS estimator is consistent” should read “the estimation method using OLS is consistently estimated a parameter.”

5. Asymptotic = Approximated. Asymptotic theory = approximation property. We are interested in how an approximation works as \( n \to \infty \).

**Estimation in the Finite Sample**

1. Unbiased: An estimator of a parameter \( \theta \) is unbiased if the mean of its sampling distribution is \( \theta \).

\[ E\left(\hat{\theta} - \theta\right) = 0 \text{ for all } n. \]

2. Efficient: An unbiased estimator \( \hat{\theta}_1 \) is more efficient than another unbiased estimator \( \hat{\theta}_2 \) is the sampling variance of \( \hat{\theta}_1 \) is less than that of \( \hat{\theta}_2 \).

\[ V\left(\hat{\theta}_1\right) < V\left(\hat{\theta}_2\right) \]
3. Mean Squared Error:

\[ \text{MSE} \left( \hat{\theta} \right) = E \left[ (\hat{\theta} - \theta)^2 \right] \]
\[ = E \left[ (\hat{\theta} - E\hat{\theta} + E\hat{\theta} - \theta)^2 \right] \]
\[ = V(\hat{\theta}) + E \left[ (\hat{\theta} - \theta) \right]^2 \]

4. Likelihood Function: rewrite

\[ u_i = y_i - bx_i \]

and consider the joint density of \( u_i \). If \( u_i \) are independent, then

\[ f (u_1, ..., u_n|b) = f (u_1|b) f (u_2|b) ... f (u_n|b) \]
\[ = \prod_{i=1}^{n} f (u_i|b) = L (b|x_1, ..., x_n) \]

The function \( L (b|\mathbf{u}) \) is called the likelihood function for \( b \) given the data \( \mathbf{u} \).

5. Cramer-Rao Lower Bound: Under regularity condition, the variance of an unbiased estimator of a parameter \( \theta \) will always be at least as large as

\[ [I(\theta)]^{-1} = \left( -E \left[ \frac{\partial^2 \ln L(\theta)}{\partial \theta^2} \right] \right)^{-1} = \left( E \left[ \frac{\partial \ln L(\theta)}{\partial \theta} \right]^2 \right)^{-1} \]

where the quantity \( I(\theta) \) is the information number for the sample.

4 Large Sample Distribution Theory

Definition and Theorem (Consistency and Convergence in Probability)

1. Convergence in probability: The random variable \( x_n \) converges in probability to a constant \( c \) if

\[ \lim_{n \to \infty} \text{Prob} (|x_n - c| > \varepsilon) = 0 \text{ for any positive } \varepsilon. \]

We denote

\[ x_n \to_p c, \text{ or plim}_{n \to \infty} x_n = c \]
Carefully look at the subscript ‘n’. This means $x_n$ is dependent on the size of $n$. For an example, the sample mean, $n^{-1}\sum_{i=1}^{n} x_i$ is a function of $n$.

2. Almost sure convergence:
   \[ \text{Prob} \left( \lim_{n \to \infty} x_n = c \right) = 1 \]
   Note that almost sure convergence is stronger than convergence in probability. We denote
   \[ x_n \xrightarrow{a.s.} c \]

3. Convergence in the $r$–th mean
   \[ \lim_{n \to \infty} E \left( |x_n - c|^r \right) = 0 \]
   and denote it as
   \[ x_n \xrightarrow{L^r} c \]
   When $r = 2$, we say convergence in quadratic mean.

4. Consistent Estimator: An estimator $\hat{\theta}_n$ of a parameter $\theta$ is a consistent estimator of $\theta$ iff
   \[ \text{plim}_{n \to \infty} \hat{\theta}_n = \theta \]

5. Khinchine’s weak law of large number: If $x_i$ is a random sample from a distribution with finite mean $E(x_i) = \mu$, then
   \[ \text{plim}_{n \to \infty} \bar{x}_n = \text{plim}_{n \to \infty} n^{-1} \sum_{i=1}^{n} x_i = \mu \]

6. Chebychev’s weak law of large number: If $x_i$ is a sample of observations such that
   \[ E(x_i) = \mu_i < \infty, V(x_i) = \sigma_i^2 < \infty, \sigma_n^2/n = n^{-2} \sum_{i=1}^{n} \sigma_i^2 \to 0 \text{ as } n \to \infty, \]
   then
   \[ \text{plim}_{n \to \infty} (\bar{x}_n - \bar{\mu}_n) = 0 \]
   where $\bar{\mu}_n = n^{-1} \sum_{i=1}^{n} \mu_i$.  

29
7. Kolmogorov’s Strong LLN: If $x_i$ is a sequence of independently distributed random variables such that $E(x_i) = \mu_i < \infty$ and $V(x_i) = \sigma_i^2 < \infty$ such that $\sum_{s=1}^{\infty} \sigma_s^2/s^2 < \infty$ as $n \to \infty$ then

$$\bar{x}_n - \bar{\mu}_n \to a.s. 0$$

8. (Corollary of 7) If $x_i$ is a sequence of iid variables such that $E(x_i) = \mu < \infty$, and $E|x_i| < \infty$, then

$$\bar{x}_n - \mu \to a.s. 0$$

9. Markov’s Strong LLN: If $x_i$ is a sequence of independent random variables with $E(x_i) = \mu_i < \infty$ and if for some $\delta > 0$, $\sum_{i=1}^{\infty} E\left[|x_i - \mu_i|^{1+\delta}\right]/i^{1+\delta} < \infty$, then

$$\bar{x}_n - \bar{\mu}_n \to a.s. 0$$

Properties of Probability Limits

1. If $x_n$ and $y_n$ are random variables with $\text{plim} x_n = b$ and $\text{plim} y_n = c$, then

$$\text{plim} (x_n + y_n) = b + c$$

$$\text{plim} x_n y_n = bc$$

$$\text{plim} \frac{x_n}{y_n} = \frac{b}{c} \text{ if } c \neq 0$$

2. $W_n$ is a matrix whose elements are random variables and if $\text{plim} W_n = \Omega$, then

$$\text{plim} W_n^{-1} = \Omega^{-1}$$

3. If $X_n$ and $Y_n$ are random matrices with $\text{plim} X_n = B$ and $\text{plim} Y_n = C$, then

$$\text{plim} X_n Y_n = BC$$
Convergence in Distribution

1. $x_n$ converges in distribution to a random variable $x$ with cdf $F(x)$ if

$$\lim_{n \to \infty} |F(x_n) - F(x)| = 0 \text{ at all continuity points of } F(x)$$

In this case, $F(x)$ is the limiting distribution of $x_n$, and this is written

$$x_n \to^d x$$

2. Cramer-Wold Device: If $x_n \to^d x$, then

$$c'x_n \to c'x$$

where $c \in \mathbb{R}$

3. Lindeberg-Levy CLT (Central limit theorem): If $x_1, \ldots, x_n$ are a random sample from a probability distribution with finite mean $\mu$ and finite variance $\sigma^2$, then it sample mean, $\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$ have the following limiting distribution

$$\sqrt{n} (\bar{x}_n - \mu) \to^d N (0, \sigma^2)$$

4. Lindegerg-Feller CLT: Suppose that $x_1, \ldots, x_n$ are a random sample from a probability distribution with finite mean $\mu_i$ and finite variance $\sigma_i^2$. Let

$$\bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2, \quad \bar{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \mu_i$$

where $\lim_{n \to \infty} \max (\sigma_i) / (n \sigma_n) = 0$. Further assume that $\lim_{n \to \infty} \bar{\sigma}_n^2 = \bar{\sigma}^2 < \infty$, then it sample mean, $\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$ have the following limiting distribution

$$\sqrt{n} (\bar{x}_n - \bar{\mu}_n) \to^d N (0, \bar{\sigma}^2)$$

or

$$\frac{\sqrt{n} (\bar{x}_n - \bar{\mu}_n)}{\bar{\sigma}} \to^d N (0, 1)$$
5. Liapounov CLT: Suppose that \( \{x_i\} \) is a sequence of independent random variables with finite mean \( \mu_i \) and finite positive variance \( \sigma_i^2 \) such that \( E(|x_i - \mu_i|^{2+\delta}) < \infty \) for some \( \delta > 0 \). If \( \sigma_n \) is positive and finite for all \( n \) sufficiently large, then

\[
\frac{\sqrt{n}(\bar{x}_n - \bar{\mu}_n)}{\sigma_n} \to^d N(0, 1)
\]

6. Multivariate Lindeberg-Feller CLT:

\[
\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \to^d N(0, \mathbf{Q})
\]

where \( V(x_i) = \mathbf{Q} \) and we assume that \( \lim \bar{Q}_n = \mathbf{Q} \)

7. Asymptotic Covariance Matrix: Suppose that

\[
\sqrt{n}(\hat{\theta}_n - \theta) \to^d N(0, \mathbf{V})
\]

then its asymptotic covariance matrix is defined as

\[
\text{Asy. Var}(\hat{\theta}_n) = \frac{1}{n} \mathbf{V}
\]

Order of A Sequence

1. A sequence \( c_n \) is of order \( n^\delta \), denoted \( O(n^\delta) \), iff

\[
\text{plim}_{n \to \infty} \frac{c_n}{n^\delta} = c < \infty
\]

(a) \( c_n = 1 = O(1) \)

(b) \( c_n = n^2 = O(n^2) \), \( \text{plim} n^2/n^2 = 1 \).

(c) \( c_n = 1/(n+10) = O(n^{-1}) \), \( \text{plim}(n+10)^{-1}/n^{-1} = 1 \)

2. A sequence \( c_n \) is of order less than \( n^\delta \) iff

\[
\text{plim}_{n \to \infty} \frac{c_n}{n^\delta} = 0
\]

(a) \( c_n = 0 = o(1) \), \( \text{plim}0/1 = 0 \).

(b) \( c_n = O(n^{-1/2}) \), then \( c_n = o(1) \)
Order in Probability

1. A sequence random variable $x_n$ is $O_p(g(n))$ if there exists some $N_\varepsilon$ such that $\varepsilon > 0$ and all $n > N_\varepsilon$,

$$\text{Prob} \left( \left| \frac{f_n}{g(n)} \right| < c \right) > 1 - \varepsilon$$

where $c$ is a finite constant

(a) If $x_n \sim N(0, \sigma^2)$, then $x_n = O_p(1)$. Since given $\varepsilon$, there is always some $c$ such that

$$\text{Prob} \left( |x_n| < c \right) > 1 - \varepsilon$$

(b) $O_p(n^a)O_p(n^b) = O_p(n^{a+b})$

(c) If $\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \rightarrow^d N(0, \sigma^2)$, then $(\bar{x}_n - \bar{\mu}_n) = O_p(n^{-1/2})$ but $\bar{x}_n = O_p(1)$

2. The notation $x_n = o_p(g_n)$ means

$$\frac{x_n}{g_n} \rightarrow^p 0$$

(a) If $\sqrt{n}\bar{x}_n \rightarrow^d N(0, \sigma^2)$, then $\bar{x}_n = O_p(n^{-1/2})$ and $\bar{x}_n = o_p(1)$

(b) $o_p(n^a)o_p(n^b) = o_p^{a+b}$
Sample Questions

Part I: Consider the following model

\[ M_1 : y_i = bx_i + u_i, \quad i = 1, \ldots, n \]
\[ M_2 : y_i = a + bx_i + u_i \]

Suppose that

\[ E x_i u_1 = c < \infty \quad \text{but} \quad E x_i u_i = 0 \text{ for all } i \]

Q1: Show the OLS estimator \( \hat{b} \) in M1 is unbiased and consistent
Q2: Show the OLS estimator \( \hat{b} \) is biased but consistent
Q3: Suppose that \( u_i \sim iidN(0, 1) \). Derive the limiting distribution of \( \hat{b} \) in M1
Q4: Suppose that \( u_i \sim iidN(0, \sigma_i^2) \). Derive the limiting distribution of \( \hat{b} \) in M2

Part II: Consider the following model

\[ y = Xb + u \]

Q5: Obtain the limiting distribution of \( \theta = x' u \)
Q6: Obtain the limiting distribution of \( \hat{b} \)
Q7: Suppose that \( u_i \sim N(0, \sigma_i^2) \). Find the asymptotic variance of \( \hat{b} \)