Lognormal Distribution

M Pal

Introduction

- The lognormal (LN) distribution is occasionally referred to as the **Galton distribution** or **Galton's distribution**, after Francis Galton. The log-normal distribution has also been associated with other names, such as McAlister, Gibrat and Cobb–Douglas.
- This distribution does not satisfy the Weak Pareto Law.
- Since it has been seen that most income data satisfy Weak Pareto Law, the LN distribution is not expected to give a good fit to the income data in the upper tail.
- A **log-normal (or lognormal) distribution** is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then Y = Ln(X) has a normal distribution. Equivalently, if Y has a normal distribution, then the exponential function of Y, X = exp(Y), has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values. It is a convenient and useful model for measurements in exact and engineering sciences, as well as medicine, economics and other topics (e.g., energies, concentrations, lengths, financial returns and other metrics).

Lognormal Distribution Defined

- Symbolically, $X \sim LN(\mu, \sigma^2)$ if $Ln(X) \sim N(\mu, \sigma^2)$, $X > 0$. In other words, X is said to be lognormally distributed with parameters μ and σ^2 if Ln(X) is normally distributed with the same parameters μ and σ^2. Thus, E(Ln(X)) = μ and V(Ln(X)) = σ^2. So, μ and σ^2 are not mean and variances of X.

- **Properties of LN Distribution**

  1. **Distribution Function (CDF):**

- Let $\Lambda(x|\mu, \sigma^2)$ denote the distribution function of LN distribution. Then

$$\Lambda(x|\mu, \sigma^2) = P(X \le x) = P(Ln(X) \le Ln(x))$$

$$= P\left(\frac{Ln(X) - \mu}{\sigma} \le \frac{Ln(x) - \mu}{\sigma}\right) = \Phi\left(\frac{Ln(x) - \mu}{\sigma}\right).$$

Properties of LN Distribution (Continued)

2. Density function (pdf):

$$\frac{d\Lambda(x|\mu,\sigma^2)}{dx} = \frac{d\Phi\left(\frac{Ln(x)-\mu}{\sigma}\right)}{dx}$$

$$= \frac{d\Phi\left(\frac{Ln(x)-\mu}{\sigma}\right)}{dLn(x)}\frac{dLn(x)}{dx} = \frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{(Ln(x)-\mu)^2}{2\sigma^2}}.$$

$$\therefore \text{pdf} = \lambda(x) = \frac{1}{\sqrt{2\pi}\sigma x}e^{-\frac{(Ln(x)-\mu)^2}{2\sigma^2}}; x > 0.$$

Properties of LN Distribution (Continued)

3. Mode:

$$\frac{d\mathrm{Ln}\big(\lambda(x)\big)}{dx} = -\frac{1}{x} - \frac{2(\mathrm{Ln}(x)-\mu)^2}{2\sigma^2}\frac{1}{x}.$$

$$\frac{d\mathrm{Ln}\big(\lambda(x)\big)}{dx}\ \Big|_{x=x_m} = 0 \Rightarrow \frac{\mathrm{Ln}(x_m)-\mu}{\sigma^2} = -1$$

$$\text{or } \mathrm{Ln}(x_m) = \mu - \sigma^2 \text{ or } x_m = e^{\mu-\sigma^2}.$$

- Verify that

$$\frac{d^2\mathrm{Ln}\big(\lambda(x)\big)}{dx^2} = \frac{1}{x^2\sigma^2}\left[\sigma^2 - 1 + \mathrm{Ln}(x) - \mu\right]$$

$$\frac{d\mathrm{Ln}\big(\lambda(x)\big)}{dx}\ \Big|_{\mathrm{Ln}(x)=\mu-\sigma^2} = -\frac{1}{(x^2\sigma^2)} < 0.$$

Properties of LN Distribution (Continued)

4. Moments:

$$E\left(X^j\right) = E\left(e^{j\,Ln(X)}\right) = E\left(e^{jY}\right);\text{ where } Y = Ln(X) \sim N(\mu, \sigma^2)$$

$$= e^{j\mu + \frac{1}{2}j^2\sigma^2},\ \text{(Recalling expression of mgf of normal distribution.)}$$

$$\text{Mean} = E(X) = e^{\mu + \frac{1}{2}\sigma^2} \text{ and } E(X^2) = e^{2\mu + 2\sigma^2}.$$

$$V(X) = e^{2\mu + 2\sigma^2} - \left(e^{\mu + \frac{1}{2}\sigma^2}\right)^2 = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right).$$

$$CV(X) = \frac{\sqrt{V(X)}}{E(X)} = \sqrt{e^{\sigma^2} - 1} = \eta, \text{say}.$$

- **Home Task 1:** Prove that

$$\text{Coefficient of Skewness} = \gamma_1 = \frac{\mu_3}{(\mu_2)^{3/2}} = \eta^3 + 3\eta > 0.$$

$$\text{Coefficient of Kurtosis} = \gamma_2 = \frac{\mu_4}{(\mu_2)^2} - 3 = \eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 > 0.$$

- Thus, the lognormal distribution is limited to non-negative values, unimodal, bell-shaped and positively skewed. However, as $\sigma \to 0$, the distribution becomes nearly symmetrical.

Properties of LN Distribution (Continued)

5. Quantiles:

- Let X_P be the Pth quantile of LN distribution $\Lambda(\mu, \sigma^2)$. Then,

$$P = \Lambda(X_P|\mu, \sigma^2) = \Phi\left(\frac{Ln(X_P) - \mu}{\sigma}\right).$$

$$or\ \frac{Ln(X_P) - \mu}{\sigma} = \Phi^{-1}(P)$$

$$or\ Ln(X_P) = \mu + \sigma\Phi^{-1}(P) = \mu + \sigma t_P, say.$$

- t_P, the normit of P, is defined by $\Phi(t_P) = P$.
- In other words, t_P is the quantile of order P of N(0, 1).
- Thus, the Pth quantile of LN distribution is

$$X_P = e^{\mu + \sigma t_P}.$$

- ***Special Case:*** Median = $X_{0.5} = e^{\mu}, since\ t_{0.5} = 0$.

Discussions

- **Note:** Empirically it was observed that "Mean – Mode = 3(Mean – Median)". But, here

$$\text{Ln(Mean)} - \text{Ln(Mode)} = 3\big(\text{Ln(Mean)} - \text{Ln(Median)}\big).$$

$$Q_3(X) = e^{\mu+0.6745\sigma} \text{ and } Q_1(X) = e^{\mu-0.6745\sigma}.$$

- Note that

$$\text{"Probable error of } \bar{x} = 0.6745\sigma/\sqrt{n}.$$

Graphical Test of Lognormality

- Graphical Test for Lognormality is based on

$$\mathrm{Ln}(X_P) = \mu + \sigma\, t_P \text{ or } t_P = \frac{\mathrm{Ln}(X_P) - \mu}{\sigma}.$$

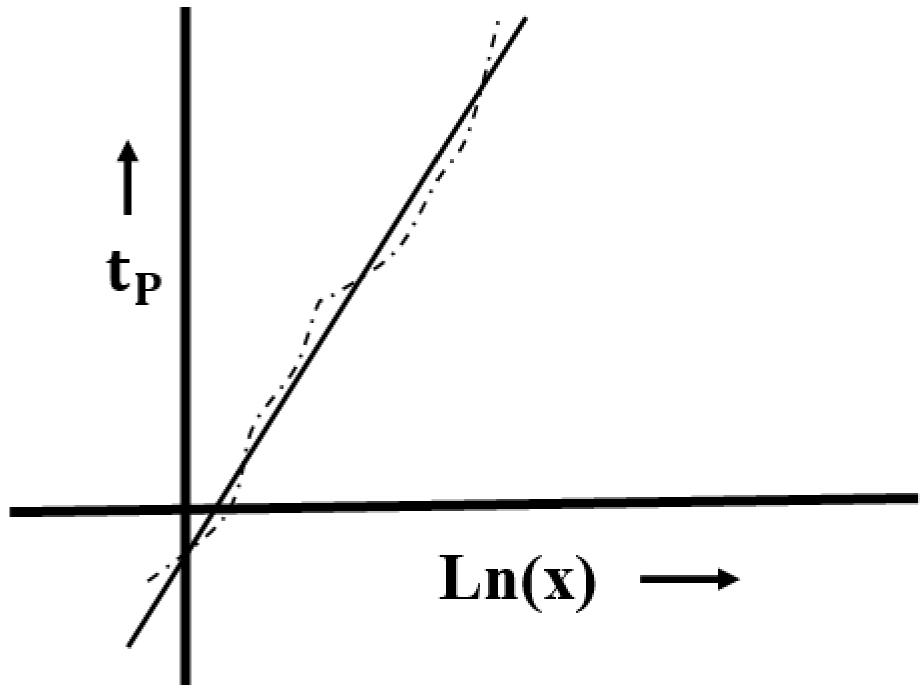
Income Class $\left(\mathbf{x_{j-1} - x_j}\right)$	**Relative Frequency** $(\mathbf{p_j})$	**Cum. Rel. frequency** $(\mathbf{P_j})$	**Log Upper boundary** $\left(\mathbf{Ln(x_j)}\right)$	**Normit** $\left(\mathbf{t_{P_j}}\right)$
$x_0 - x_1$	p_1	$P_1 = p_1$	$\mathrm{Ln}(x_1)$	t_{P_1}
$x_1 - x_2$	p_2	$P_2 = p_1 + p_2$	$\mathrm{Ln}(x_2)$	t_{P_2}
---	---	---	---	---
$x_{j-1} - x_j$	p_i	$P_j = \sum_{i=1}^{j} p_i$	$\mathrm{Ln}(x_j)$	t_{P_j}
---	---	---	---	---
$x_{K-1} - x_K$	p_K	$P_K = \sum_{i=1}^{K} p_i = 1$	$\mathrm{Ln}(x_K)$	t_{P_K}
Total	1	---	---	---

Graphical Test of Lognormality (Continued)

- To examine whether the underlying distribution can be taken as lognormal apart from sampling errors, let x_j be the quantile of order P_j and so

$$\mathrm{Ln}(x_j) \approx \mu + \sigma t_{P_j} \text{ or } t_{P_j} = \frac{\mathrm{Ln}(x_j) - \mu}{\sigma}.$$

- So, we plot t_{P_j} against $\mathrm{Ln}(x_j)$ examine subjectively whether the graph is reasonably straight.



Graphical Test of Lognormality (Continued)

- If the graph is straight, the underlying distribution is lognormal (LN).

$$t_{P_j} = \frac{Ln(x_j) - \mu}{\sigma} = -\frac{\mu}{\sigma} + \frac{1}{\sigma} Ln(x_j).$$

$$\therefore Slope = \frac{1}{\sigma}, and\ Intercept = -\frac{\mu}{\sigma}.$$

- If one passes a line by judgement, one gets quick estimates of μ and σ.

Estimation of Parameters in the LN Distribution

- **Estimation of Parameters in the LN Distribution :**
- **Raw Data**

1. We have a random sample of the observations $x_1, x_2, \dots x_n$ from $LN(\mu, \sigma^2)$.

- Note that $Ln(x_1), Ln(x_2), \dots, Ln(x_n)$ form a random sample from $N(\mu, \sigma^2)$. Hence,

$$\tilde{\mu} = \frac{1}{n}\sum Ln(x_j)\,, and\ \tilde{\sigma}^2 = \frac{1}{n}\sum\left(Ln(x_j) - \tilde{\mu}\right)^2$$

- are ML estimates of μ and σ^2. However, we take

$$\hat{\mu} = \tilde{\mu} = \frac{1}{n}\sum Ln(x_j)\,, and\ \hat{\sigma}^2 = \frac{1}{n-1}\sum\left(Ln(x_j) - \hat{\mu}\right)^2,$$

- because $\hat{\sigma}^2$ is unbiased estimate for σ^2.

Estimation (Continued)

- **Method of Moments' estimators:** Note that the usual Method of Moment estimator for μ and σ^2 can be obtained from the first two moments as

$$\mu = Ln\big(E(X)\big) - \frac{1}{2}\left(1 + \frac{V(X)}{\big(E(X)\big)^2}\right)$$

$$\sigma^2 = Ln\left(1 + \frac{V(X)}{\big(E(X)\big)^2}\right)$$

- Method of Moments' estimators are very much inefficient.

Estimation (Continued)

- **Grouped Data**
- We have the following grouped data.

Income Class $(x_{j-1} - x_j)$	Frequency (n_j)
$x_0 - x_1$	n_1
$x_1 - x_2$	n_2
---	---
$x_{j-1} - x_j$	n_j
---	---
$x_{K-1} - x_K$	n_K
Total	n

Estimation (Continued)

- We assume that the observations are independent random sample from $LN(\mu, \sigma^2)$. It follows that $n_1, n_2, \dots n_K$ follows multinomial distribution with probability mass function (pmf)

$$P(n_1, n_2, \dots n_K) = \frac{n!}{n_1!\, n_2! \dots n_K!} \prod_{1}^{K} \left(\pi_j\right)^{n_j},$$

- where,

$$\pi_j = \Lambda\left(x_j | \mu, \sigma^2\right) - \Lambda\left(x_{j-1} | \mu, \sigma^2\right)$$

$$= \Phi\left(\frac{Ln\left(x_j\right) - \mu}{\sigma}\right) - \Phi\left(\frac{Ln\left(x_{j-1}\right) - \mu}{\sigma}\right).$$

- The likelihood function L is same as the pmf, but viewed as function of the unknown parameters μ and σ^2.

$$Ln(L) = Const. + \sum_{1}^{K} n_j Ln\left(\pi_j(\mu, \sigma^2)\right).$$

- The likelihood function is numerically maximized to get ML estimates.

Estimation (Continued)

$$\frac{\partial Ln(L)}{\partial \mu} = -\frac{1}{\sigma}\sum\left[n_j \frac{\varphi\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\varphi\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}{\Phi\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\Phi\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}\right]$$

$$\frac{\partial Ln(L)}{\partial \sigma}$$

$$= -\frac{1}{\sigma}\sum\left[n_j \frac{\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)\varphi\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)\varphi\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}{\Phi\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\Phi\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}\right]$$

$$= \frac{1}{\sigma}\sum\left[n_j \frac{\varphi'\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\varphi'\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}{\Phi\left(\frac{Ln\left(x_j\right)-\mu}{\sigma}\right)-\Phi\left(\frac{Ln\left(x_{j-1}\right)-\mu}{\sigma}\right)}\right],$$

- since $\varphi'(x) = \frac{d}{dx}\left(\frac{1}{\sqrt{2\pi}}e^{\frac{x^2}{2}}\right) = -x\varphi(x)$.
- We put $\frac{\partial Ln(L)}{\partial \mu} = 0 \; and \; \frac{\partial Ln(L)}{\partial \sigma} = 0$, and solve by numerical methods.

Short-Cut Methods Which are Reasonably Efficient

- ML methods lead to cumbersome equations which are solved by numerical methods. However, if the class intervals are equal on log scale (e.g., 1-2, 2-4, 4-8, …) then method of moments estimators are almost as efficient as MLEs. I.e., With the grouped data in log scale one can compute mean and standard deviation of Ln(x) values by usual methods – applying Sheppard's correction if necessary and this method of moment estimators are almost as efficient as MLEs.
- **Short-Cut Methods Which are Reasonably Efficient**
- 1. **Method of Quantiles:**
- Here we equate two chosen quantiles, say, X_{P_1} and X_{P_2} to the theoretical expressions for the corresponding true quantiles and solve the equations for μ and σ^2.

$$X_{P_1} = e^{(\mu+\sigma t_{P_1})} and\ X_{P_2} = e^{(\mu+\sigma t_{P_2})}.$$

- or, $$Ln(X_{P_1}) = (\mu + \sigma t_{P_1})\ and\ Ln(X_{P_2}) = (\mu + \sigma t_{P_2}).$$

Short-Cut Methods Which are Reasonably Efficient (Continued)

- It was found that maximum efficiency will be attained if $P_1 \approx 1 - P_2$.
- Note that

$$\text{Efficiency } of\ quatile\ estimate$$

$$= \frac{Asymptotic\ variance\ of\ MLE}{Asymptotic\ variance\ of\ quatile\ estimate} = Eff(p), say.$$

- For estimation of μ, efficiency is maximum if we use $X_{0.27}$ and $X_{0.73}$. Then efficiency is about 80%. For estimation of σ, efficiency is maximum if we use $X_{0.07}$ and $X_{0.93}$. Then efficiency is about 65%.

Fitting of LN Distribution

- **Fitting of LN Distribution:**
- This is described in the following table. We assume that we have already estimated μ and σ^2 as $\hat{\mu}$ and $\hat{\sigma}^2$.

Income Class $(x_{j-1} - x_j)$	Frequency (n_j)	$t_j = \frac{Ln(x_j) - \hat{\mu}}{\hat{\sigma}}$	$\Phi(t_j)$	$\Delta\Phi(t_j)$	$n\Delta\Phi(t_j)$
$0 - x_0$	0				
$x_0 - x_1$	n_1				
$x_1 - x_2$	n_2				
---	---				
$x_{j-1} - x_j$	n_i				
---	---				
$x_{K-1} - x_K$	n_K				
$x_K - \infty$	0				
Total	n				

Moment Distribution of LN Distribution

- **Moment Distribution of LN Distribution:**
- Let X be a non-negative continuous random variable with distribution function F(.). Consider

$$F_j(x) = \frac{\int_0^x u^j dF(u)}{\int_0^\infty u^j dF(u)}, j = 1, 2, 3, \ldots$$

- Formally $F_j(x)$ has all the properties of a distribution function. The distribution having this as its distribution function (for a particular value of j) is referred to as the jth moment distribution of the random variable X. F_1, the first moment distribution, is needed in connection with Lorenz Curve, F_3 is needed in small particle statistics. Diameter of these particles are lognormally distributed.

The Moment Distribution of LN Distribution

- **Theorem:** The jth moment distribution associated with $\Lambda(\mu, \sigma^2)$ is also a LN distribution. Precisely,

$$\Lambda_j(x|\mu, \sigma^2) = \Lambda(x|\mu + j\sigma^2, \sigma^2).$$

- Proof:

$$\Lambda_j(x|\mu, \sigma^2) = \frac{\int_0^x u^j d\Lambda(u|\mu, \sigma^2)}{\int_0^\infty u^j d\Lambda(u|\mu, \sigma^2)}$$

$$= \frac{\int_0^x e^{jLn(u)} d\Lambda(u|\mu, \sigma^2)}{e^{j\mu + j^2\sigma^2/2}} = A\int_0^x e^{jLn(u)} d\Lambda(u|\mu, \sigma^2)\,, say,$$

- where

$$A = e^{-j\mu - j^2\sigma^2/2}$$

The Moment Distribution of LN Distribution (Continued)

$$= A\int_0^x \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2\sigma^2}\left\{(Ln(u))^2 - 2\mu Ln(u) + \mu^2 - 2\sigma^2 jLn(u)\right\}} du$$

$$= A\int_0^x \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2\sigma^2}\left\{(Ln(u))^2 - 2\left(\mu + j\sigma^2\right)Ln(u) + \left(\mu + j\sigma^2\right)^2 - j^2\sigma^4 - 2\mu j\sigma^2\right\}} du$$

$$= A\int_0^x \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2\sigma^2}\left(Ln(u) - \left(\mu + j\sigma^2\right)\right)^2 + j\mu + \frac{j^2\sigma^2}{2}} du$$

$$= \int_0^x \frac{1}{\sqrt{2\pi}\sigma u} e^{-\frac{1}{2\sigma^2}\left\{Ln(u) - \left(\mu + j\sigma^2\right)\right\}^2} du$$

$$= \Lambda(x|\mu + j\sigma^2, \sigma^2). \qquad\qquad QED$$

LC of LN Distribution

$$X \sim \Lambda(\mu, \sigma^2) \Rightarrow F(x) = P(X \leq x) = \Phi\left(\frac{Ln(x) - \mu}{\sigma}\right). \qquad \ldots(I)$$

$$F_1(x) = \Lambda_1(x|\mu, \sigma^2) = \Lambda(x|\mu + \sigma^2, \sigma^2)$$

- (by moment distribution property)

$$= \Phi\left(\frac{Ln(x) - (\mu + \sigma^2)}{\sigma}\right) = \Phi\left(\frac{Ln(x) - \mu}{\sigma} - \sigma\right) \qquad \ldots(II)$$

- From (I), $t_F = \frac{Ln(x) - \mu}{\sigma}$.
- From (II), $t_{F_1} = \frac{Ln(x) - \mu}{\sigma} - \sigma$.
- Hence $t_{F_1} = t_F - \sigma$.
- This is the equation of Lorenz Curve of $\Lambda(\mu, \sigma^2)$.

LR of LN Distribution

$$LR = 1 - 2\int_0^1 F_1 dF = 1 - 2\int_0^1 \Lambda_1 d\Lambda$$

$$= 1 - 2\int_0^1 \Lambda_1(\mathrm{x}|\mu, \sigma^2) d\Lambda(\mathrm{x}|\mu, \sigma^2)$$

$$= 1 - 2\int_0^1 \Lambda(\mathrm{x}|\mu + \sigma^2, \sigma^2) d\Lambda(\mathrm{x}|\mu, \sigma^2)$$

$$= 1 - 2I, say.$$

- To evaluate I consider two independent random variables X_1 and X_2 such that

$$X_1 \sim \Lambda(\mu, \sigma^2) \; and \; X_2 \sim \Lambda(\mu + \sigma^2, \sigma^2).$$

- Observe that I = P($X_2 \leq X_1$).

LR of LN Distribution (Continued)

$$P(X_2 \leq X_1) = P\left(\frac{X_2}{X_1} \leq 1\right)$$

$$= P(\mathrm{Ln}(X_2) - \mathrm{Ln}(X_1) \leq 0)$$

$$= \Phi\left(\frac{0 - \sigma^2}{\sqrt{2\sigma^2}}\right), \text{since } \mathrm{Ln}(X_2) - \mathrm{Ln}(X_1) \sim N(\sigma^2, 2\sigma^2)$$

- Thus,

$$P(X_2 \leq X_1) = \Phi\left(-\frac{\sigma}{\sqrt{2}}\right) = 1 - \Phi\left(\frac{\sigma}{\sqrt{2}}\right).$$

- Hence,

$$LR = 1 - 2\left[1 - \Phi\left(\frac{\sigma}{\sqrt{2}}\right)\right] = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1.$$

Properties of LC of LN Distribution

- **1. Property 1:**

$$t_{F_1} = t_F - \sigma$$

- The curve depends on σ and not on μ. As σ increases F_1 decreases for a given F. I.e., between two LN distributions, the one with greater σ has a uniformly lower LC. In other words, LCs of two LN distributions (with different σ) never intersects except at the end points.
- As σ → 0, LC → egalitarian line. Also, as σ → 0, F_1→F.
- As σ → ∞, F_1 → 0, whenever F < 1, case of perfect inequality.
- Any reasonable measure of inequality will be a monotone increasing function of σ. S.D.(log) measure is particularly meaningful when distribution is LN.

Properties of LC of LN Distribution (Continued)

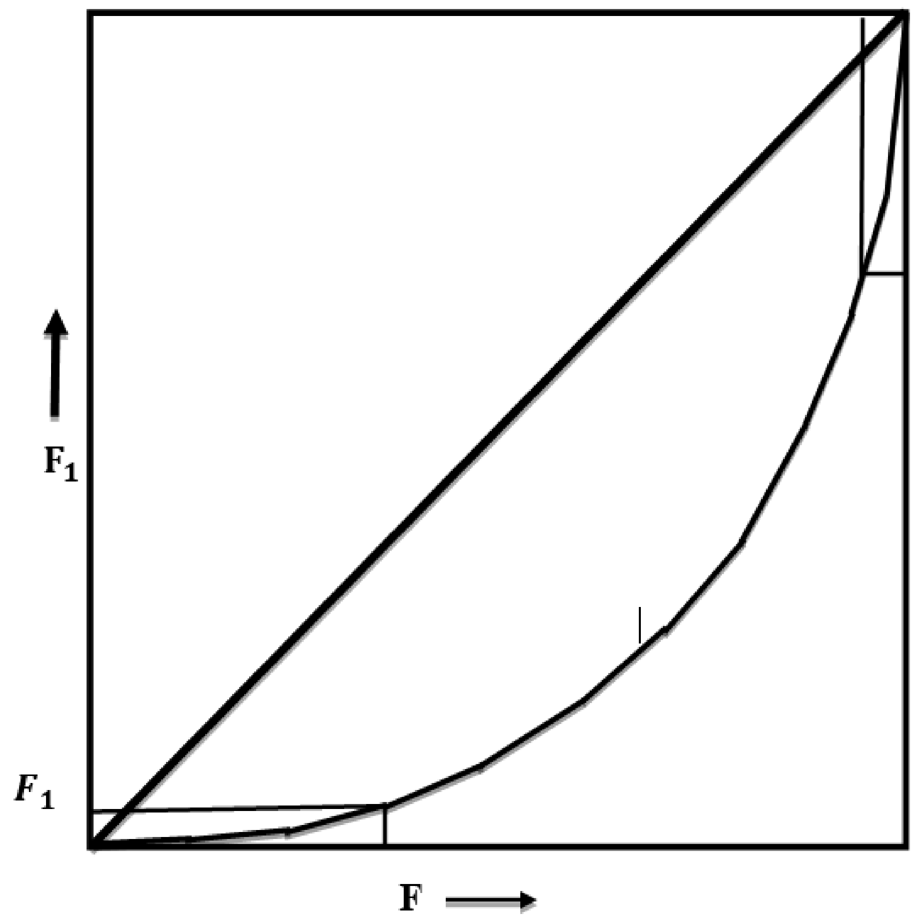
- **2. Property 2 (Symmetry):**

$$t_{F_1} = t_F - \sigma$$

$$\Rightarrow -t_F = -t_{F_1} - \sigma$$

$$\Rightarrow t_{1-F} = t_{1-F_1} - \sigma.$$

- So, if (F, F_1) is a point on the LC of the LN distribution then $(1 - F_1, 1 - F)$ is also a point on the same curve. If bottom 10% get a share of 3%, then top 3% get a share of 10%. So, the LC is symmetrical about the diagonal $F_1 + F = 1$.



Properties of LC of LN Distribution (Continued)

3. **Property 3:**
   - The Lorenz Curve intersects the diagonal at F + F_1 =1 orthogonally and the underlying value of X is the mean of the distribution, i.e., $e^{\mu+\sigma^2/2}$.
   - **Proof:** $F(x) = \Phi\left(\frac{Ln(x)-\mu}{\sigma}\right).$
   - Put $x = e^{\mu+\sigma^2/2} = M, say.$
   - Observe that $F(M) = \Phi\left(\frac{\sigma}{2}\right).$
   - Similarly, $F_1(M) = \Phi\left(-\frac{\sigma}{2}\right) = 1 - \Phi\left(\frac{\sigma}{2}\right),$
   - *since* $F_1(x) = \Phi\left(\frac{Ln(x)-\mu}{\sigma} - \sigma\right).$
   - *So,* $F(M) + F_1(M) = 1. and$
   - $\frac{dF_1}{dF} = \frac{x}{mean} \Rightarrow \frac{dF_1}{dF}\Big|_{x=mean} = 1.$ $Q.E.D.$

Thank You