Lecture - 3

Dr. Arabin Kumar Dey

Assistant Professor
Department of Mathematics
Indian Institute of Technology Guwahati

July 31, 2013

Outline

Outline

- A random variable is a numerical quantity whose value cannot be predicted or measured with certainty.
- Example : Change in stock price is random when it can take on many possible values.

Random Variable

Suppose the function $X : \Omega \rightarrow \mathcal{R}$ is defined on the probability space $(\Omega, \mathcal{F}, P)$. X is said to be a real-valued $\mathcal{F}$- measurable random variable if for each real number x, we have

$$\{w : X(w) \leq x\} \in \mathcal{F}$$

Example

Let $\Omega = \{H, T\}$ and $\rho =$ class of all subset of Ω.
Define X by $X(H) = 1$, $X(T) = 0$

$$X^{-1}(-\infty, x] = \begin{cases} \phi \in \rho & if \quad x < 0 \\ \{T\} \in \rho & if \quad 0 \leq x < 1 \\ \{H, T\} \in \rho & if \quad x \geq 1 \end{cases}$$

Therefore, X is an r.v.

Example

X is not a random variable

$$\Omega = \{a, b, c\}, \qquad \text{Å} = \big\{\{a, b, c\}, \{a, b\}, \{c\}, \phi\big\}$$

define three random variables X, Y, Z as follows;

W	X	Y	Z
a	1	1	1
b	1	2	7
c	2	2	4

Which of the random variables X, Y, Z are Å-measurable.

X is Å-measurable because $\{X \leq 1\} = \{a, b\} \in$ Å
and $\{X \leq 2\} = \{a, b, c\} \in$ Å
Y is not r.v. as $\{Y \leq 1\} = \{a\} \notin$ Å.
Z is not r.v. as $\{Z \leq 1\} = \{a\} \notin$ Å

Discrete and Continuous random variable

- Assume that each time step the price can either increase or decrease by a fixed amount $\Delta > 0$. Suppose, P_1, where $0 < P_1 < 1$ is the probability of an increase and $P_2 = 1 - P_1$ is the probability of a decrease.

Discrete and Continuous random variable

- Assume that each time step the price can either increase or decrease by a fixed amount $\Delta > 0$. Suppose, P_1, where $0 < P_1 < 1$ is the probability of an increase and $P_2 = 1 - P_1$ is the probability of a decrease.
- A random variable such as this one with only a finite number of values is called discrete.

Discrete and Continuous random variable

- Assume that each time step the price can either increase or decrease by a fixed amount $\Delta > 0$. Suppose, P_1, where $0 < P_1 < 1$ is the probability of an increase and $P_2 = 1 - P_1$ is the probability of a decrease.
- A random variable such as this one with only a finite number of values is called discrete.
- This discrete random variable can take infinite sequence of numbers.

Discrete and Continuous random variable

- Assume that each time step the price can either increase or decrease by a fixed amount $\Delta > 0$. Suppose, P_1, where $0 < P_1 < 1$ is the probability of an increase and $P_2 = 1 - P_1$ is the probability of a decrease.
- A random variable such as this one with only a finite number of values is called discrete.
- This discrete random variable can take infinite sequence of numbers.
- For example let N be the total number of steps in random walk until a down-step. Assume steps are independent and the probability that $N = n$ is $P_1^{n-1}P_2$, which the probability of $(n-1)$ up-steps and then a down-step.

Discrete and Continuous random variable

- Assume that each time step the price can either increase or decrease by a fixed amount $\Delta > 0$. Suppose, P_1, where $0 < P_1 < 1$ is the probability of an increase and $P_2 = 1 - P_1$ is the probability of a decrease.
- A random variable such as this one with only a finite number of values is called discrete.
- This discrete random variable can take infinite sequence of numbers.
- For example let N be the total number of steps in random walk until a down-step. Assume steps are independent and the probability that $N = n$ is $P_1^{n-1}P_2$, which the probability of $(n-1)$ up-steps and then a down-step.
- If we consider N as number of increases of stock price until a down-step. In that case probability that $N = n$ is $P_1^n P_2$. Clearly N can take any value in the sequence $\{0, 1, 2, \cdots,\}$. These are two forms of well-known geometric distribution.

Continuous Random Variable

Definition

A random variable X taking on any value in some interval is called Continuous random variable. For example, loss of a bank can take any value in some interval or can take any real number.

A continuous random variable has a probability density function (PDF), often called simply a density, f_X such that

$$P[X \in A] = \int_A f_X(x)dx$$

Cumulative distribution function

The cumulative distribution function(CDF) of X is defined as

$$F_X(x) = P[X \leq x]$$

If X has a PDF f_X then

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

The CDF is often more useful for finding probabilities than the PDF. For example suppose that we want the probability that X is between a and b for some pair of numbers $a < b$. For simplicity assume that X is continuous so that the probability of X equaling exactly a or exactly b is zero.

Characterization of random variable

We write $P[a \leq X \leq b] = F_X(b) - F_X(a)$ which is easily calculated assuming that F_X can be evaluated.

Note that: For discrete distribution, $F_X(x)$ is defined as

$$F_X(x) = \sum_{z:z \leq x} P(X = z).$$

- $F_X(x)$ is continuous random variable, then X is continuous random variable.
- $F_X(x)$ discrete random variable, then X is discrete random variable.
- $F_X(x)$ mixture random variable, then X is said to be mixture random variable.

Outline

1. Random Variable and distribution function
2. Continuous Probability Distributions
3. Some Theoretical Properties of Distribution :
4. Heavy tail Distributions
   - Mixture Models
5. Quantitle/Percentile:

The density of the uniform distribution on the interval $[a, b]$ is given by the formula :

$$f_{a,b}(x) = \begin{cases} \frac{1}{b-a} & if \quad a \leq x \leq b \\ 0 & \quad otherwise \end{cases}$$

The corresponding cumulative distribution function is given by :

$$F_{a,b}(x) = \begin{cases} 0 & if \quad x < a \\ \frac{x-a}{b-a} & if \quad a \leq x \leq b \\ 1 & if \quad x > b \end{cases}$$

The probability distribution is denoted by $U(a, b)$. The uniform distribution over the unit interval $[0, 1]$ is frequently used.

Result: If U follows $U[0,1]$, and $V = a + (b-a)U$ then V follows $U(a,b)$.
Proof: $P(V \leq x) = P(a + (b-a)U \leq x) = P(U \leq \frac{x-a}{b-a}) = \frac{x-a}{b-a}$ for $a \leq x \leq b$. Clearly this implies that V follows uniform distribution.

- Location paramter is parameter that **shifts a distribution to the right or left** without changing the distribution's shape or standard deviation. If $f(x)$ is any fixed density, then $f(x-\mu)$ is a family of distributions with location parameter μ.
- A parameter is a scale parameter if it is a constant multiple of the standard deviation. $\theta^{-1}f(x/\theta), \quad \theta > 0$ is a family of distributions with a scale parameter θ.
- A shape parameter generally refers to any parameter that is neither location nor scale parameter. This parameter is responsible for the shape of the p.d.f. or p.m.f.. For exmaple the parameter p of Binomial(n, p) distribution is shape parameter.

Definition

Normal Distribution: The univariate normal distribution, also called the Gaussian distribution, is most often defined by means of its density function.

It depends upon two parameters μ and σ^2, and is given by the formula :

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}; \quad x \in R \quad \mu \in R \quad \sigma > 0.$$

The two parameter μ and σ^2 are the mean and variance of the distribution respectively.

- Indeed, if X is a random variable with such a distribution (in which case we use the notation $X \sim N(\mu, \sigma^2)$ we have : $E(X) = \mu$ and $V(X) = \sigma^2$.
- The corresponding c.d.f.

$$F_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} f_{\mu,\sigma^2}(x)dx$$

cannot be given by a formula in closed form involving standard function.

- For this we introduce a special notation Φ_{μ,σ^2} for this function.
- When $\mu = 0$ and $\sigma^2 = 1$ we call it a **standard normal distribution** and we use the notation for this probability distribution as $N(0, 1)$.

$$X \sim N(\mu, \sigma^2) \Leftrightarrow Z = \left(\frac{X-\mu}{\sigma}\right) \sim N(0,1)$$

Because of this fact, most computations are done with the $N(0,1)$ distribution only.

- $P[-\sigma \le X-\mu \le \sigma] = P[-1 \le Z \le 1] = \Phi(1) - \Phi(-1) = 0.683$
- $P[-2\sigma \le X-\mu \le 2\sigma] = P[-2 \le Z \le 2] = \Phi(2) - \Phi(-2) = 0.955$
- $P[-3\sigma \le X-\mu \le 3\sigma] = P[-3 \le Z \le 3] = 0.997$

These facts can be restated in words on :

- The probability that a normal r.v. is one standard deviation, or less away from its mean is 0.683
- The probability that normal r.v. is two standard deviation, or less away from its mean is 0.955
- The probability that normal r.v. is three standard deviation, or less away from its mean is 0.997

Definition

Log-Normal Distribution: Y is log normal if $X = \ln Y$ is normal. If $X \sim N(\mu, \sigma^2)$, then $Y = \exp(X)$ is log-normal with mean and variance

$$E(Y) = e^{\mu+\sigma^2/2}, \qquad V(Y) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$

Conversely if Y is lognormal with mean μ_y and variance σ_y^2, then $X = \ln Y$ is normal with mean and variance

$$E(X) = \ln\left[\frac{\mu_y}{\sqrt{1 + \frac{\sigma_y^2}{\mu_y^2}}}\right]$$

$$V(X) = \ln\left[1 + \frac{\sigma_y^2}{\mu_y^2}\right]$$

Application : If the log-return of an asset is normally distributed with mean 0.0119 and standard deviation 0.0663, Then what is the mean and standard deviation of its simple return ?

Answer : Solve this problem in two steps.

Step 1 : Based on the prior results, the mean and variance of $Y_t = \exp(r_t)$ are

$$E(Y) = \exp\left(0.0119 + \frac{0.0663^2}{2}\right) = 1.014$$

$$V(Y) = \exp(2 \times 0.0119 + 0.0663^2)[\exp(0.0663^2) - 1] = 0.0045$$

Step 2 : Simple return is $R_t = \exp(r_t) - 1 = (Y_t - 1)$. Therefore,
$E(R) = E(Y) - 1 = 0.014$
$V(R) = V(Y) = 0.0045$, Standard deviation $= \sqrt{V(R)} = 0.067$.

Definition

The Exponential Distribution: An Extremely useful distribution, particularly in internet traffic, insurance, catastrophe and rainfall modeling, failure or reliability problem.

The density function can be given by

$$f_\lambda(x) = \begin{cases} 0 & if \quad x < 0 \\ \lambda e^{-\lambda x} & if \quad x \geq 0 \end{cases}$$

The positive number λ is called the rate of the distribution the corresponding c.d.f is given by the following formula :

$$F_\lambda(x) = \begin{cases} 0 & if \quad x < 0 \\ (1 - e^{-\lambda x}) & if \quad x \geq 0 \end{cases}$$

Definition

Weibull Distribution: The probability density function can be written as

$$f_{WE}(x;\beta,\theta) = \begin{cases} \beta\theta^{\beta}x^{\beta-1}e^{-(\theta x)^{\beta}}; & x>0,\ \theta>0,\ \beta>0 \\ 0 & \text{otherwise} \end{cases}$$

Here, θ is scale parameter and β determines the shape parameter. For $\beta < 1$, shape parameter of weibull decays sharply like exponential distribution whereas $\beta > 1$ we find Weibull as bell-shaped. For $\beta = 2$ we get Rayleigh distribution.

Definition

Beta Distribution: This distribution is defined on [0,1]. The p.d.f. can be given by

$$f(x;\alpha_1,\alpha_2) = \begin{cases} \frac{1}{B(\alpha_1,\alpha_2)}x^{\alpha_1-1}(1-x)^{\alpha_2-1}; & 0<x<1 \\ 0 & \text{otherwise} \end{cases}$$

where

Outline

- **Moments of a random variable X with density $f(x)$:**

  l-th moment

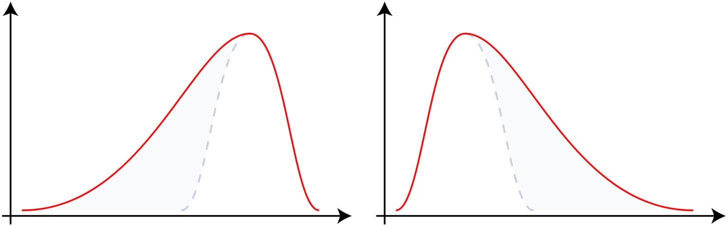
  $$m_l^1 = E(X^l) = \int_{-\infty}^{\infty} x^l f(x) dx$$

  First moment : mean or expectation of X

- **l-th central moment**

  $$m_l = E((X - \mu_x)^l) = \int_{-\infty}^{\infty} (x - \mu_x)^l f(x) dx$$

- **2nd central moment :** Variance of X.



Negative Skew   Positive Skew

- **Skewness(symmetry)**   Skewness is mathematically defined as .

$$S(x) = E\left[\frac{(X-\mu_x)^3}{\sigma_x^3}\right]$$

- It measures the symmetry of a distribution.
- If X is symmetric about μ_x. Then $S(x) = 0$.
- In finance when we consider the distribution of return series negative skewness means there is a substantial probability of big negative return. Similarly, positive skewness means there is a greater than normal probability of a big positive return.

Kurtosis (fat-fail) Mathematically, Kurtosis is defined by the quantity :

$$K(x) = E\left[\frac{(X-\mu_x)}{\sigma_x}\right]^4$$

- Note that for normal distribution Kurtosis is 3.
- We define excess kurtosis as $(K(x) - 3)$.
- High Kurtosis implies heavy (or long) tails in distribution with respect to normal distribution.
- Negative numbers of excess kurtosis indicates platykurtic (thinner tail) distribution whereas positive numbers indicate a leptokurtic (fatter tail) distribution
- In finance, a measure of the fatness of the tails of kurtosin where there is large losses on an investment i.e. excess kurtosis indiactes that the volatility of the investment is itself highly volatile or volatility of volatility.

Outline

Heavy tail Distributions

Definition

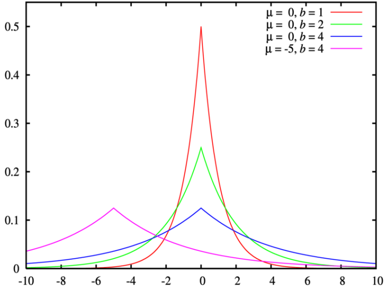
Distribution with high tail probabilities compared to a reference distribution (say normal) with same mean and standard deviation are called heavy-tailed.

- Since kurtosis is more sensitive to tail-weight, high kurtosis is nearly (missing something) with having heavy tails. A heavy tailed distribution is more prone to extreme values, which are sometime called outliers.
- In financial application one is especially concerned when the return distribution has heavy tails because of the possibility of an extremely large negative return which could entirely deplete the capital reveres of a firm.

Examples :

Double exponential distribution Double exponential has slightly heavier tails than normal distributions. Its pdf can be given as :

$$f(x|\mu, b) = \frac{1}{2b} e^{-\left(\frac{|x-\mu|}{b}\right)} \qquad \mu \in R, b > 0$$



- This fact can be appreciates by comparing their densities, the density of double exponential with scale parameter θ is proportional to $\exp(-|\frac{x}{\theta}|)$
- The density of the $N(0, \sigma^2)$ distribution is proportional to $\exp(-0.5(\frac{x}{\sigma})^2)$.
- The term $-x^2$ converges to $-\infty$ much faster than $-|x|$ as $|x| \longrightarrow \infty$.
- Therefore the normal density converges to 0 much faster than the double exponential density as $x \longrightarrow \infty$.

- No density of the form

$$\exp\left(-\left|\frac{x}{\theta}\right|^{\alpha}\right) \tag{1}$$

(where $0 < \alpha < 2$ is a shape parameter and θ scale parameter) will have truly heavy tails, and in particular $E|X|^n < \infty$ for all n. No matter how large, and in addition there is finite mean and variance.

- To achieve very heavy tails, the density must behave like $|x|^{-(a+1)}$ for some $a > 0$, which we call polynomial tails.

Pareto Distribution : Swiss economics professor *Vilfredo Pareto (1848 - 1923)* formulates an eponymous law which states that the function of a population with income exceeding an amount x is equal to

$$cx^{-a} \qquad for \quad all \quad x \tag{2}$$

Where c and a are positive constant independent of x but depending on the population. If $F(x)$ is the CDF of the income distribution, then (2) implies that $F(x) = 1 - \left(\frac{c}{x}\right)^{a}; \quad x > c$. where, c is the minimum income.

Therefore the p.d.f. corresponding to above type is

$$f(x) = \frac{ac^a}{x^{a+1}}, \quad x > c$$

So a Pareto distribution has polynomial tails. As mentioned before constant a is called the tail-index. It is also called the Pareto constant
Distribution with Pareto tail : A cumulative distribution function $F(x)$ is said to have **Pareto right tail** if its survival function satisfies

$$(1 - F(x)) = L(x)x^{-a}$$

For some $a > 0$ where $L(x)$ is slowly varying function.

A function is said to be slowly varying at ∞ if

$$\frac{L(tx)}{L(x)} \longrightarrow 1 \quad as \quad x \longrightarrow \infty \quad \text{for} \quad t > 0.$$

For example $\log x$ is slowly varying as

$$\frac{\log(tx)}{\log(x)} = \frac{\log t + \log x}{\log x} \longrightarrow 1 \quad x \longrightarrow \infty \quad \text{for} \quad t > 0.$$

t-distribution have heavy tails

t-distribution's density is proportional to

$$\frac{1}{\left[1+\left(\frac{x^2}{\nu}\right)\right]^{\frac{\nu+1}{2}}}$$

which for large values of $|x|$ is approximately

$$\frac{1}{\left[1+\left(\frac{x^2}{\nu}\right)\right]^{\frac{\nu+1}{2}}} \propto \frac{1}{|x|^{-(\nu+1)}}$$

Therefore, t-distribution has polynomial tails with $a = \nu$. Any distribution with polynomial tails has heavy tails, and the smaller value of ν the heavier the tails. From modeling perspective, a problem with the t-distribution is that the tail index is **integer valued**, rather than **a continuous parameter**, which limits the flexibility of the t-distribution as a model for financial market data.

Definition

The Cauchy Distribution: The density function can be given by

$$f_{m,\lambda}(x) = \frac{1}{\pi}\frac{\lambda}{[\lambda^2 + (x-m)^2]}, \quad x \in R$$

The distribution is denoted by $c(m, \lambda)$.

$$F_{m,\lambda}(x) = \int_{-\infty}^{x} f_{m,\lambda}(x)dx = \frac{1}{\pi}\tan^{-1}\frac{(x-m)}{\lambda} + \frac{1}{2}$$

The mean of cauchy distribution does not exist as

$$\int_{-\infty}^{\infty} \frac{\lambda|x|dx}{\pi[\lambda^2 + (x-m)^2]^2}$$

does not exist. Therefore, the higher order moments also do not exist. Tail of cauchy is heavier than normal.

Outline

Mixture Models

Another class of heavy tailed distribution is mixture models.

- Consider a distribution which is 90% $N(0,1)$ and 10% $N(0,25)$.
- This is an example of a normal mixture distribution since it is a mixture of two different normal distributions called the components.
- The variance of this distribution is $(0.9)(1)+(0.1)(25)=3.4$ so its standard deviation is $\sqrt{3.4}=1.84$ **This distribution is much different than an** $N(0,3.4)$ distribution, **even though both distributions have same mean and variance.**

- In general p.d.f. can be written as

$$f(x) = pf_1(x;\theta_1) + (1-p)f_2(x;\theta_2)$$

Where $f_1(x;\theta_1)$ and $f_2(x;\theta_2)$ are two different p.d.f. let $F_1(x;\theta_1)$ and $F_2(x;\theta_2)$ are two corresponding c.d.f's.

- $F(x) = pF_1(x;\theta_1) + (1-p)F_2(x;\theta_2)$.
- For a $N(0,\sigma^2)$ random variable X,

$$P\Big[|X| > x\Big] = 2\Big[1 - \Phi\Big(\frac{x}{\sigma}\Big)\Big]$$

- Therefore, for the normal distribution with variance 3.4,

$$P_N\left[|X| > 6\right] = 2\left[1 - \Phi\left(\frac{6}{\sqrt{3.4}}\right)\right] = 0.0011$$

- For the normal mixture population which has variance 1 with probability 0.9 and variance 25 with probability 0.1 we have that

$$\begin{aligned} P_{MN}\left[|X| > 6\right] &= 2\Big[0.9(1 - \Phi(6)) + 0.1(1 - \Phi(6/5))\Big] \\ &= 2[0.9(0) + (0.1)(0.115)] = 0.023 \end{aligned}$$

- Since, $\frac{0.023}{0.0011} \approx 21$, **the normal mixture distribution is 21 times more likely to be in this outlier range than the $N(0, 3.4)$ population**, even though both have a variance of 3.4.

It is not difficult to compute the Kurtosis of this mixture.

- If $Z \sim N(\mu, \sigma^2)$, then $E(Z-\mu)^4 = 3\sigma^4$. Therefore if X has normal mixture distribution. then

$$E\Big((X-\mu)^4\Big) = 3(0.9 + (0.1)25^2) = 190.2\,; \boxed{\text{here } \mu = 0}$$

- and Kurtosis of X is $\frac{190.2}{3.4^2} = 16.453$, which is clearly very high indicating the heavy tail behavior.
- **Despite the heavy tails of a normal mixture, the tails are exponential, not polynomial**.

Outline

- If the CDF of X is continuous and strictly increasing then it has an inverse function F^{-1}. For each q between 0 and 1, $F^{-1}(q)$ is called the **q-quantile or 100qth percentile**.
- The probability that a continuous X is below its q-quantile is precisely q, but we will show this is not exactly true for discrete random variables.
- The median is 50% percentile or 0.5-quantile. The 25% and 75% percentiles (0.25 and 0.75-quantiles) are called the first and third quartiles and the median is second quartile. The three quartiles divide the range of the range of the random variable into four groups of equal probability.
- Similarly the 20%, 40%, 60% and 80% percentiles are called quintiles and 10%, 20%, $\cdots$, 90% percentiles are called deciles.

- For discrete distributon we may see $P(X < F^{-1}(q)) < q < P(X \leq F^{-1}(q))$.
- Also if the function is increasing but not strictly increasing, there exists an interval of q-quantiles. Therefore we modify the definition of quantile : *The set q-quantile is the closed interval* $[x_q^-, x_q^+]$ *where:* $x_q^- = \inf\{x : F(x) \geq q\}$ *and* $x_q^+ = \inf\{x : F(x) > q\}$.

Quantile-Quantile plot

Example

For exponential distribution qth qunatile π_q can be written as

$$\pi_q = \frac{1}{\lambda} \ln(\frac{1}{1-q})$$

whereas for cauchy it can be written as

$$\pi_q = m + \lambda \tan(q\pi - \frac{\pi}{2})$$

	$X \sim N(0,1)$	$X \sim C(0,1)$
$\pi_{0.8} = F_X^{-1}(0.8)$	0.842	1.376
$\pi_{0.85} = F_X^{-1}(0.85)$	1.036	1.963
$\pi_{0.9} = F_X^{-1}(0.9)$	1.282	3.078
$\pi_{0.95} = F_X^{-1}(0.95)$	1.645	6.314
$\pi_{0.975} = F_X^{-1}(0.975)$	1.960	12.706
$\pi_{0.99} = F_X^{-1}(0.975)$	2.326	31.821

In particular, we would have to plot the points
$(0.842,\ 1.376),\ (1.036,\ 1.963),\ (1.282,\ 3.078),\ \cdots,\ (2.326,\ 31.821)$.

- Note that all these points are above diagonal $y = x$, and in fact they drift further and further away above the diagonal.
- This fact is at the core of the interpretation of a QQ-plot comparing two distributions: points above the diagonal in the rightmost part of the plot indicate that the upper tail of the first distribution (whose quantiles are on the horizontal axis) is thinner than the tail of the distribution whose quantiles are on the verticle axis.
- Similarly, points below the diagonal on the left part of the plot indicate that the second distribution has a heavier lower tail.