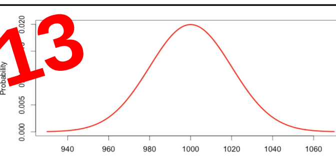
Concepts	Description
Common probability distribution	
Probability distribution Probability function	Probabilities of all possible outcomes for a random variable. Sum of all probabilities of all possible outcomes = 1 Probability function : probability that a random variable = a specific value p(x) = P(X=x)
Discrete random variable vs. Continuous random variable	Discrete randome variable - Limited number of possible outcomes - A measurable and positive probabilities for each outcome - p(x) = 0 → x cannot occur, p(x) > 0 → x can occur Continuous random variable - Unlimited number of possible outcomes - Only measurable probabilities for a range of outcome. Probability of any single outcome = 0 - Could only consider P(x1 ≤ X ≤ x2); p(x1) = p(x2) = 0
Cummulative distribution function (cdf)	Cummulative distribution function : probability that a random variable ≤ a specific values F(x) = P(X ≤ x)
Discrete uniform random variable	Discrete uniform random variable : probabilities for all possible outcomes are equal p(x1) = p(x2) = p(x3) = p(x4) = p(x5)
Binomial random variable	Binomial random variable : number of "success" in a given number of trials, where outcome can be either "success" or "failure" Bernoulli random variable : Binominal random variable in which the number of trials = 1 $p(x) = \frac{n!}{(n-x)! \times x!} \times p^x \times (1-p)^{n-x}$ In which: $p = probability\ of\ success\ in\ each\ trial`$
Expected value and variance of Binomial random variable	Expected value of X = E(X) = n × p Variance of X = n × p × (1 - p)
Tracking error	Tracking error = total return of the portfolio - total return of the benchmark
Continuous uniform distribution	Continuous uniform distribution : where probability of X occuring in a possible range = Length of the range ÷ total of all possible values $a \le x_1 \le x_2 \le\le b \rightarrow P(x_1 \le X \le x_2) = \frac{x_2 - x_1}{b - a}$
Normal distribution	Characteristic of normal distribution : - Symetrical, bell-shaped, single peak (@ the exact center of the distribution) - Mean = Median = Mode (@ the exact center of the distribution) - Skew = 0 - Kurtosis = 3 
Univariate distributions / Multivariate distributions	Univariate distributions : Distribution of a single random variable Multivariate distributions : probabilities for more than 1 variables, meaningful when behaviour of each random variable is dependent upon the behavour of others Multivariate could be describe by : - mean of individual random variables $(\mu_1, \mu_2, \mu_3, \ldots, \mu_n)$ - variance of each individual randome variables $(\sigma_1{}^2, \sigma_2{}^2, \sigma_3{}^2, \ldots, \sigma_n{}^2)$ - pair wise correlation $= 0.5 \times n \times (n-1)$
Confident interval	Confident interval : range of values around an expected outcome within which we expect the actual outcome to be some specific % of the time $\bar{X} - 1s \le X \le \bar{X} + 1s \rightarrow 68\%\ confidence\ interval$ $\bar{X} - 1.65s \le X \le \bar{X} + 1.65s \rightarrow 90\%\ confidence\ interval$ $\bar{X} - 1.96s \le X \le \bar{X} + 1.96s \rightarrow 95\%\ confidence\ interval$ $\bar{X} - 2.58s \le X \le \bar{X} + 2.58s \rightarrow 99\%\ confidence\ interval$
Standard normal distribution / z-value	Standard normal distribution : has mean = 0 ; standard deviation = 1 z-value : number of standard deviations a given observationis from the population mean. $z - value = \frac{observation\ - population\ mean}{standard\ deviation} = \frac{x - \mu}{\sigma}$
Shortfall risk	Shortfall risk : probability that a portfolio's value < a specific value over a given period of time. ↑ Safety first ratio → more preferable $Safety\ First\ Ratio = \frac{E(R_P) - R_T}{\sigma_P}$ In which: $R_T = Target\ return$
Lognormal distribution	$Lognormal\ distribution = e^x\ , where\ x\ is\ normally\ distributed$ Characteristic of lognormal distribution : - Lognormal distribution is skewed to the right - Lognormal distribution is bounded from below by zero Lognormal distribution is often used to model asset prices (because it cannot be negative, and can take any positive value)

Concepts	Description
Sampling and estimation	
Simple random sampling	Simple random sampling : method of selecting a sample in such a way that each item / person in the population being studied has the same likelihood of being included in the sample
Sampling distribution	Sampling distribution : probability **distribution of all possible sample statistics** compounded from a set of equal equal-size samples that were randomly drawn from the sample population
Sampling error	Sampling error : difference between sample statistic (mean, variance, standard deviation of the sample) and its corresponding population parameter (mean, variance, standard deviation of the population) E.g.: Sampling error of mean = sample mean - population mean
Stratified random sampling	Stratified random sampling : use a classification system to separate the population → smaller groups, based on 1 or more distinguishing characteristics
Time-series data / Cross-sectional data / Longtitudinal data / Panel data	Time-series data : consist observations taken *over a period of time* at specific and equally spaced time intervals (e.g.: monthly returns of Stock A from 2014 to 2017) Cross-sectional data : sample of observation taken *at a single point in time* (e.g.: EPS of all Stock as at 31/12/2017) Longtitudinal data : observations over time of *multiple characteristics of same entity* (e.g.: GDP, inflation, unemployment rate of Vietnam from 2014 to 2017) Panel data : observation over time of *same characteristic for multipled entities* (e.g.: EPS of all companies for the recent 3 years)
Central limit theorem	Central limit theorem : for simple random samples size n (from a population with mean µ, finite variance σ^2), sampling distribution of the sample mean ($\bar{x}$) approaches a normal probability distribution with mean µ and variance = $\frac{\sigma^2}{n}$ as the sample size becomes larger Sample size (n) is sufficiently large (n≥ 30) → distribution of sample means will be approximately normal Mean of the population (µ) = mean of the distribution of all possible sample means
Standard error of the sample mean	Standard error of the sample mean : standard deviation of the distribution of the sample means $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ in which : $\sigma_{\bar{x}} = standard\ error\ of\ the\ sample\ mean$ $\sigma = standard\ deviation\ of\ the\ population$ $n = sample\ size$ In case the standard deviation of the population is unknown, could use the standard deviaton of the sample $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ in which : $s_{\bar{x}} = standard\ error\ of\ the\ sample\ mean$ $s = standard\ deviation\ of\ the\ sample$ $n = sample\ size$
Desirable properties of an estimator	Desirable properties of an estimator : - Unbiasedness : sign of estimation error is random - Efficiency : lower sampling error than any other unbiased estimator - Consistency : variance of sampling error decreases with sample size
Point estimates / Confidential internal	Point estimates : sample values used to estimate population parameter (e.g.: sample mean is an estimator of the population mean) Confidential interval : range of values in which the population parameter is expected to lie
Student's t-distribution	Student's t-distribution is : - bell-shaped probability distribution - symmetrical about its mean (mean = 0) - to construct confidence intervals based on small samples (n < 30) from populations with unknown variance and a normal distribution - defined by a single parameter : degree of fredom (df) = number of sample observations (n) - 1 (for sample mean) - more probability in the tails than normal distribution - ↑ df →more observations near the center of the distribution + ↓ % of observations in tails → shape of t-distribution more closely approaches a standard normal distribution
Confidence interval	Confidence interval : range of values within which the actual value of parameter will lie, given the probability of 1 - α In which: α = level of significant 1 - α = degree of confidence Formula for confidence interval for the population mean - normal distribution with a known variance $Confidence\ interval = Point\ estimate \pm Reliability\ factor\ \times standard\ error = \bar{x} \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ Formula for confidence inerval for the population mean - normal distribution with unknown variance $Confidence\ interval = \bar{x} \pm t_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}$ Examples of interpretation: - Probabilistic interpretation : 99% of resulting confidence intervals will include the population mean - Practical interpretation : 99% confident that the population mean score is between 73.55 and 86.45

Criteria for selecting the appropriate test statistic		Small sample (n < 30)	Large sample (n > 30)
	Normal distribution - known variance	z - statistic	z - statistic
	Normal distribution - unknown variance	t - statistic	t - statistic
	Nonnormal distribution - known variance	N/A	z - statistic
	Nonnormal distribution - unknown variance	N/A	t - statistic
	(Note : samples are drawn randomly from the population)		