**Probability**

- Statistical Concepts and Market Returns
- Probability Concepts
- Common Probability Distributions

## Statistical Concepts and Market Returns

**Statistics**

- **Descriptive Statistics**
  - Summarize characteristics of a large dataset
- **Inferential Statistics**
  - Make forecasts, estimates, or judgments based on a smaller **sample** set

**Population and Sample**

- **Population**
  - The set of all possible members
- **Sample**
  - A subset of the population on interest

**Measurement Scales**

- **Nominal Scales 定类测量**
  - Categorical, no order 无序类别
  - 数学：等于、不等于
- **Ordinal Scales 定序测量**
  - Categorical, 有序类别
  - 数学：大于、小于
- **Interval Scale 定距测量**
  - Differences between scale values are equal
  - No absolute zero point
  - 数学：加减
- **Ratio Scales 定比测量**
  - Ratio between scale values are equal
  - A **true** zero point
  - 数学：乘除

**Measures – Parameter & Statistic**

- Population: **parameter 参数**
- Sample: sample **statistic 统计量**

**Frequency Distribution**

- **Step**
  - Define the intervals [x, y]
    - Mutually exclusive
  - Tally the observations
  - Count the observations

- **Measures**
    - Absolute Frequency
    - Relative Frequency
    - Cumulative absolute/relative frequency
- **Plot**
    - Histogram 直方图
        - Frequency ~ Interval
    - Frequency polygon
        - connect midpoint of each interval

**Central Tendency Measure 中心**

- Population Mean 种群平均
    - $\mu = \frac{1}{N}\sum_{i=1}^{n} X_i$, where N is the population size
- Sample Mean 样本平均
    - $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, where n is the sample size
- Arithmetic Mean 算数平均
    - Sum of the deviations from the mean is zero
- Weighted Mean 加权平均
    - $\overline{X_w} = \frac{1}{n}\sum_{i=1}^{n} w_i X_i$, where $\sum_{i=1}^{n} w_i = 1$
    - **Examples**: portfolio return
- Median 中位数
    - Midpoint of sorted data, 50% quantile
    - Practice: midpoint or the average of the two middle points of
    - Robust to outliers
- Mode 众数
    - The value that occurs most frequently
    - Can have more modes (multimodal) or no mode
    - Unimodal: one mode
    - multimodal: bimodal/trimodal
- Geometric Mean 几何平均
    - $G = \sqrt[n]{X_1 \times X_2 \cdots X_n}$
    - **Examples**: investment return over multiple periods, HPR
- Harmonic Mean 调和平均
    - $G = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}} = \frac{N}{\sum_{i=1}^{N} 1/X_i}$
    - Examples: average cost
- Mean Comparison
    - Harmonic mean < geometric mean < arithmetic mean

**Quantiles 分位数**

- **Quantile 分位数**
    - A value below which a stated partition of the data in a distribution lies
    - The location of a given percentile y: $L_y = (n+1) \times \frac{y}{100}$
- **Examples**

- Quartiles – 4 quarters
- Quintile – 5 fifths
- Decile – 10 tenths
- Percentile – 100 hundredths

**Dispersion Tendency Measure 分散**

- Dispersion is the **variability** around the **central** tendency
- Range 间距
    - $\text{Range} = \text{maximum value} - \text{minimum value}$
- Inter-Quantile Range (IQR)
    - $\text{IQR} = \text{Q}_{0.75} - Q_{0.25}$
- Mean Absolute Deviation (MAD)平均绝对偏差
    - $\text{MAD} = \frac{\sum_{i=1}^{N}|X_i - \bar{X}|}{N}$
- Population Variance 种群方差
    - $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$, where σ is population standard deviation
- Sample Variance 样本方差
    - $\text{s}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$, where s is sample standard deviation
    - Unbiased estimator
    - Dividing by n yield a biased estimator
- Chebyshev's inequity 切比雪夫不等式
    - The percentage of observations that lie within k standard deviations of the mean is at lease $1 - 1/\text{k}^2$.
    - $\text{P}\left(\frac{|X-\mu|}{\sigma} \le \text{k}\right) \ge 1 - \frac{1}{\text{k}^2}$
    - It applies to any distribution. If the distribution is normal, we can get more precise
- Relative Dispersion
    - The amount of variability in a distribution relative to a reference point or benchmark
- Coefficient of Variation (CV) 变异系数
    - Coefficient of variation is the variation per unit return $\text{CV} = \frac{s_x}{\bar{X}}$
- Sharpe Ratio 夏普率
    - $\text{Sharpe Ratio} = \frac{r_p - r_f}{s}$
        - $\text{r}_\text{f}$ is risk-free return
        - $\text{r}_\text{p} - r_f$ is excess return
    - **Limitation**
        - If two Sharpe ratios are negative, higher does not mean better
            - Higher s leads to a higher Sharpe ratio (close to zero)
        - It is useful when the distribution is symmetric.

**Distribution – Symmetry & Skewness 偏度**

- Symmetric and Asymmetric (skewed)
    - Symmetric: the shape is symmetric

- Outliers 异常值
  - Observations with extraordinarily large values
- Asymmetric **Skewness**
  - Negatively/Left skewed
    - More outliers in the lower/left tail
  - Positively/Right skewed
    - More outliers in the higher/right tail
- Relative Location
  - Symmetric: mode = median = mean
  - Positively skewed: mode < median < mean
  - Negatively skewed: mode > median > mean
- Skewness
  - $sample\ skewness = \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \bar{X})^3}{s^3}$
  - Absolute value > 0.5 indicate significant levels of skewness
- Skewness and Symmetry
  - Symmetric: = 0
  - Positively skewed: > 0
  - Negatively skewed: <0

**Distribution – Kurtosis 峰度**

- **Kurtosis**
  - The degree to which a distribution is more or less "**peaked**" than a normal distribution
  - $sample\ kurtosis = \frac{1}{n}\frac{\sum_{i=1}^{n}(X_i - \bar{X})^4}{s^4}$
  - Normal distribution = 3
- **Excess Kurtosis**
  - Normal distribution = 3
  - **Excess** kurtosis = kurtosis – 3
- **Types**
  - **Leptokurtic 高峰度**: more peaked, fat tails, excess kurtosis > 0
  - **Mesokurtic**: equal, excess kurtosis = 0
  - **Platy**kurtic 低峰度: less peaked, thin tails, excess kurtosis < 0
- **Leptokurtic – Fat Tail**
  - **More** returns **around** the mean
  - **More** returns **deviate** from the mean

**Average Return – Arithmetic & Geometric**

- **Geometric – past & future multi-year**
  - Compounding return
  - Good for measure **past** performance
  - Good for **predicting multi-year** return
- **Arithmetic – prediction next**
  - Good for predict **next** year's return

# Probability Concepts

**Random Variable and Outcome**
- **Random** Variable: an uncertain number
- **Outcome**: an observed value of a random variable

**Events**
- **Event**: a single outcome or a set of outcomes
- **Mutually** exclusive: events that cannot happen at the same time
- **Exhaustive**: all the possible events

**Probability**
- Defining Properties
  - $0 \leq \mathrm{P(E_i)} \leq 1$: positive
  - $\sum_{\mathrm{i}} \mathrm{P(E_i)} = 1$ mutual exclusive and exhaustive
- **Objective** probability
  - **Empirical** probability: established by analysing past data
  - **Priori** probability: formal reasoning and inspection process
- **Subjective** probability: least formal method, personal judgment

**Odds**
- Odds: an event will or will not occur
  - $\mathrm{Odds} = \frac{p}{1-p} = \frac{1}{n}$
  - 1-to-n => $\mathrm{p} = \frac{1}{n+1}$
- Odds against the event
  - $\mathrm{Odds\ Against} = \frac{1-p}{p} = n$
  - n-to-1

**Conditional and Joint**
- Unconditional/margin probability: P(A)
- Conditional: P(A|B), likelihood
- Joint probability: P(AB)

**Multiplication + Addition**
- Multiplication: $\mathrm{P(AB) = P(B|A)P(A) = P(A|B)P(B)}$
- Addition: $\mathrm{P(A\ or\ B) = P(A) + P(B) - P(AB)}$
- Total: $\mathrm{P(A)} = \sum_{\mathrm{X}} P(A|X)\, P(X)$

**Mutual Exclusive**
- $\mathrm{P(A|B)} = P(B|A) = 0$
- $\mathrm{P(AB)} = 0$
- $\mathrm{P(A\ or\ B)} = \mathrm{P(A)} + P(B)$

**Independent**
- $\mathrm{P(A) = P(A|B), P(B) = P(B|A)}$
- $\mathrm{P(AB) = P(A)P(B)}$

- $\mathrm{P(A\ or\ B)} = \mathrm{P(A)} + P(B) - P(A)P(B)$

**Expected**

- Expected Value
  - $\mathrm{E}(X) = \sum \mathrm{P_i} X_i$
- Variance
  - $\mathrm{Var}(X) = \sum \mathrm{P_i}(X_i - \bar{X})^2 = E(X^2) - E(X)^2$
- Covariance
  - $\mathrm{Cov}(X, Y) = \sum_\mathrm{i}\sum_\mathrm{j} \mathrm{P_{ij}}(X_i - \bar{X})(Y_j - \bar{Y}) = E(XY) - E(X)E(Y)$
- Correlation Coefficient
  - $\mathrm{Corr}(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E(X)^2}\sqrt{E(Y^2)-E(Y)^2}}$
  - Linear relationship, no units, -1 to 1
  - The **absolute** value indicates the strength of the correlation

**Portfolio**

- Define
  - For asset i, weight $\mathrm{w_i}$, and expected return $\mathrm{r_i}$
  - weight vector w, return vector r
- Expectation
  - $\mathrm{E(R_p)} = \vec{w}^\mathrm{T} \vec{r} = \sum_\mathrm{i} w_i E(R_i)$
- Variance
  - $\mathrm{Var(R_p)} = \vec{w}^\mathrm{T} \times \mathbf{\Sigma} \times \vec{w} = \sum_\mathrm{i}\sum_j w_i w_j Cov(R_i, R_j)$
- Covariance between assets
  - $\mathrm{Cov(R_i, R_j)} = \rho(\mathrm{R_i, R_j})\sigma(R_i)\sigma(R_j) = \rho_{ij}\sigma_i\sigma_j$
- Variance Matrix $\mathbf{\Sigma}$
  - $\Sigma_\mathrm{ij} = \mathrm{Cov(R_i, R_j)} = \rho_{ij}\sigma_i\sigma_j$
- Correlation and Covariance Matrix $\mathbf{Corr}$
  - $\Sigma = \bar{\sigma}^\mathrm{T} \times \mathrm{Corr} \times \bar{\sigma}$
- Two Asset Portfolio
  - $\mathrm{E(R_p)} = w_1 R_1 + w_2 R_2$
  - $\mathrm{Var(R_p)} = \mathrm{w}_1^2\sigma_1^2 + 2w_1 w_2 \sigma_1\sigma_2 + w_2^2\sigma_2^2$

**Bayes' Formula** - **Tree Diagram**

- Bayes
  - $\mathrm{P(B|A)} = \frac{P(A|B)}{P(A)} P(B) = \frac{P(A|B)}{\sum_X P(A|X)P(X)} \times P(B)$
  - $\mathrm{Posterior} = \frac{probability\ of\ likelihood}{unconditional\ probability\ of\ new\ information} \times prior$

**Counting**

- Factorial
  - $\mathrm{n! = n \times (n-1) \times \cdots 2 \times 1}$
- Permutation (order is important)
  - $A_n^k = \frac{n!}{(n-k)!}$
- Combination (order is not important)

- Binomial $\mathrm{C}_{\mathrm{n}}^{\mathrm{k}}=\frac{\mathrm{A}_{\mathrm{n}}^{k}}{k!}=\frac{n!}{(n-k)!k!}$
- Labelling
  - n items each can receive k different labels
  - multinomial $\mathrm{L}_{\mathrm{n}}^{\mathrm{k}}=\frac{n!}{\prod_{i} n_{i}!}$
    - binominal is the special case when n=2
  - $n=\sum_{i} n_{i}$
- Multiplication rule
  - K steps required to complete a task, $\prod_{i=1}^{k} \mathrm{n}_{\mathrm{i}}$

## Common Probability Distributions

### Probability Distribution

- Random Variable $X$ 随机变量
  - Discrete Random Variable 离散: the number of events can be **counted**
  - Continuous Random Variable 连续: the number of events is **infinite**
- Probability distribution 概率分布
  - The probabilities of all possible outcomes
- **Probability Function** 概率函数
  - $\mathrm{p(x)} = P(X = x)$, the probability of a random variable is equal to a specific value
  - Discrete
    - Probability **mass** function (pmf) 概率质量函数
    - $\mathrm{p(x)} = 0$ means it cannot occur
  - Continuous
    - Probability **density** function (pdf) 概率密度函数
    - $\mathrm{p(x)} = 0$ for all x even they can occur, point probability is zero
    - $\mathrm{P(x_1 \leq X \leq x_2) = P(x_1 < X < x_2)}$, range probability
  - In finance, a discrete distribution can be treated as continuous distribution if the number of outcomes is large
- Cumulative Probability Distribution 累积概率分布
- **Cumulative Distribution Function** 累积概率函数/分布函数
  - $\mathrm{F(x) = P(X \leq x)}$, the probability that $\mathrm{X}$ takes a value less than or equal to a specific value $\mathrm{x}$
  - $\mathrm{P(x_1 < X \leq x_2) = F(x_2) - F(x_1)}$

### Common Discrete Distribution

- Uniform Distribution
  - $\mathrm{p(x) = P(X = x)} = \frac{1}{\mathrm{n}}$
  - $\mathrm{F(x_i)} = \frac{\mathrm{i}}{\mathrm{n}}$
- Bernoulli Distribution
  - A trial with success probability p
  - $\mathrm{P(X = 1) = p, P(X = 0) = 1 - p}$
  - $\mathrm{Mean = p}$
  - $\mathrm{Variance = p \times (1 - p)}$
- Binomial Distribution 二项分布
  - n independent Bernoulli trials
  - $\mathrm{P(X = k) = C_n^k} p^k (1 - p)^{n-k}$
  - $\mathrm{Mean = n \times p}$
  - $\mathrm{Variance = n \times p \times (1 - p)}$
- Geometric Distribution 几何分布
- Poisson Distribution 泊松分布
  - $\mathrm{P(X = x)} = \frac{\lambda^x}{x!} e^{-\lambda}$
  - mean is $\lambda$
  - variance is $\lambda$

- unit intensity r, for time t, $\lambda = r \times t$
- Binomial and Poisson Distribution
  - When n is large, and p is small, then Binomial distribution approaches a Poisson distribution with $\lambda = n \times p$

**Uniform Continuous Distribution**

- X is uniformly distribution with range $[a, b]$
- Probability density function $f(x) = \frac{1}{b-a}$
- Distribution function $F(x) = \frac{x-a}{b-a}$
- mean is $\frac{a+b}{2}$
- variance is $\frac{(b-a)^2}{12}$

**Normal 正态分布**

- Probability Density function
  - $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-u)^2}{2\sigma^2}\right)$
- $X \sim N(\mu, \sigma^2)$
- **Properties**
  - Skewness = 0
  - Kurtosis = 3
- **Linear Properties**
  - $X \sim N(\mu, \sigma^2)$
  - $kX \sim N(k \times \mu, k^2 \times \sigma^2)$
  - $aX + b \sim N(a \times \mu + b, k^2 \times \sigma^2)$
  - $k_1X + k_2Y \sim N\left(k_1\mu_x + k_2u_y, k_1^2\sigma_x^2 + k_2^2\sigma_y^2 + 2k_1k_2\rho_{12}\sigma_1\sigma_2\right)$
- **Confidence Interval**
  - Interval: $\boldsymbol{\mu} \pm \mathbf{k} \times \boldsymbol{\sigma}$
  - 68% within 1 standard deviation
  - 95% within 2 standard deviation
  - 99% within 3 standard deviation
- **Standard Normal Distribution**
  - $Z \sim N(0,1)$
  - Standardization: $Z = \frac{X-\mu}{\sigma}$, z-score
- **Cumulative Probability**
  - $F(z) = P(X \leq z)$
  - Symmetry: $F(-z) = 1 - F(z)$
  - Upper tail: $P(X > z) = 1 - F(z) = F(-z)$
- **Roy's Safety-first Criterion**
  - The optimal portfolio minimizes the probability that the return falls below some minimum acceptable level (threshold)
    - minimize $P(R_p < R_L)$, where $R_L$ is **threshold** level return
    - $P(R_p < R_L) = P\left(\frac{R_p - E(R_p)}{\sigma_p} < \frac{R_L - E(R_p)}{\sigma_p}\right) = P\left(Z < \frac{R_L - E(R_p)}{\sigma_p}\right)$
      - $Z = \frac{R_p - E(R_p)}{\sigma_p} \sim N(0,1)$ is standard normal variable,

        - $\frac{R_L - \mathrm{E}(\mathrm{R_p})}{\sigma_p}$ is a normalized value
      - $\min \mathrm{P}(\mathrm{R_p} < R_L) \rightarrow \min \frac{R_L - \mathrm{E}(\mathrm{R_p})}{\sigma_p} \rightarrow \max \frac{\mathrm{E}(\mathrm{R_p}) - R_L}{\sigma_p}$
  - If it is normally distributed
      - $\text{maximum SFRatio} = \frac{|\mathrm{E}(\mathrm{R_p}) - R_L|}{\sigma_p}$
        - make negative return meaningful
      - Probability is $F(z)$, where $z = \frac{\mathrm{R_L} - E(R_p)}{\sigma_p} < 0$
- **Multivariate Normal**
  - Multivariate distribution: The probability with a group of random variables
  - **Correlation** Matrix

**Log-Normal Distribution 对数正态分布**

- The log of a distribution is normal, $e^x$, where x is normally distributed
- Lower bounded by zero
- Positive skewed
- Model Return Rate
  - price relative $\frac{P_2}{P_1}$
  - $\mathrm{P}_2 = P_1 e^{rt} = P_1(1 + HPR)$
  - $rt = \ln \frac{P_2}{P_1}$
  - The continuous rate is additive
- Discrete Compounding
  - Stated rate $\mathrm{R_d}$, then effective annual return (annual HPR) $\left(1 + \frac{\mathrm{R_d}}{\mathrm{m}}\right)^m - 1$
- Continuous Compounding
  - Stated rate $\mathrm{R_c}$, then effective annual return $\mathrm{e}^{\mathrm{R_c}} - 1$

**Exponential**

**Monte Carlo Simulation**

- Repeated sample from distribution
- Applications
  - Value complex securities
  - Compute VaR
  - Simulate profits/losses
  - Simulate pension fund assets and liabilities
- Limitation
  - Complex, computational expensive

**Historical Simulation**

- Based on historical data
- Limitations
  - Past may not be a good indicator of future
  - Infrequent events may be ignored