Common Probability Distribution

```
Key Words: uniform distribution, binominal distribution, cdf, normal
distribution, confidence interval, z-values, safety-first ratio, log-normal
distribution, T-distribution, chi-square distribution, F-Distribution, Monte
Carlo simulation
```

1. Uniform and Binomial Distribution

- Probability Distribution
    - discrete random variable→discrete probability distribution
    - probability function $P(x) = P(X = x)$
    - continuous random variable→continuous probability distribution
    - Difference between discrete dis(A) and continuous dis(B):
        1. $P(x) = 0$:
            - A the event not occur
            - B the event x can occur, since x is a `single point` in a range
        2. $P(x)$:
            - A: the probability that random variable $X = x$
            - B: we only consider $P(x_1 \leq X \leq x_2)$, which $= P(x_1 < X < x_2)$, since $P(x_1) = P(x_2) = 0$
- Cumulative Distribution Function(cdf)
    - $F(x) = P(X \leq x)$
    - intuition: takes on a value equal to or less than a specific value
- Discrete Uniform Random Distribution
    - probabilities for all possible outcomes for a discrete r.v. are **equal**.
    - Cumulative distribution function: $F(x_n) = np(x)$.

        - the probability of a range of outcomes is $P(x)k$,where k is the number of possible outcomes in this range.

- Properties of Continuous Uniform Distribution

    - $a \leq x_1 \leq x_2 \leq b$, where a and b are boundaries
    - $P(X < a \; or \; X > b)$=0, when X is outside the value
    - $P(x_1 \leq X \leq x_2) = \frac{x_2 - x_1}{b-a}$, the probability that outcomes locating in the range of $[x_1, x_2]$
    - Graphically, cdf of continuous uniform dist is `linear`, sum up to 1.

- Bernoulli Random Variable, Binomial random variable, Binomial Distribution

    - Binomial random variable: define the number of "success" in a given number of experiments(n), which the outcomes can be "success" or "failure".
        - p: the possibility of success
    - Bernoulli Random Variable: when $n = 1$, the special case of binominal random variable.
    - Binominal Distribution:
        - Define the probability of x success in n trials.
        - Function: $p(x) = \frac{n!}{(n-x)!x!} p^x(1-p)^{n-x} = \binom{n}{x} p^x(1-p)^{n-x}$
            - $\binom{n}{x}$:number of ways to choose x from n.
        - Expected Value and Variance:
            - the expected number of success with n trials:$E(X) = np$
            - $Var(X) = np(1-p)$

2. Normal Distribution

- Properties:

    - $X \, N(\mu, \sigma^2)$
    - $Skewness = 0$, mean=median=mode, symmetric about mean
    - $Kurtosis = 3$
    - Linear combination of normally distributed variables is also normally distributed.
        - Eg: if the return of each stock in a portfolio is normally distributed, the return on the portfolio will also follow normal distribution.

- Multivariate VS Univariate Distribution
  - Univariate Distribution:
    - the distribution of single random variable
  - Multivariate Distribution:
    - associated with multiple random variables regardless discrete or continuous variables
    - Between 2 discrete variables can be described using **joint probability tables**
  - The Role of *Correlation* in Multivariate Normal Distribution
    - one important feature distinguish multivariate dist from univariate dist
    - the strength of *linear relationship* between a pair of r.v.
    - parameterized with n assets:
      - n means: $(\mu_1, \mu_2, ..., \mu_n)$
      - n variance: $(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)$
      - pair-wised correlation: $0.5n(n-1)$
    - the lower the better, because of lower portfolio variance.

- Confidence Interval(CI)
  - Definition: a range of values around the expected outcome within which we expected the actual outcome to be specified percentage of a time.
  - For *normal distribution*, CI width depends on **expected value** and **std**. $E(X) \pm k\sigma$
    - $E(X) \pm \sigma$:68%
    - $E(X) \pm 2\sigma$:95%
  - Sometimes we need to use **sample mean and sample std** to estimate population situation:
    - 90% CI for X: $[\bar{x} - 1.65 * s, \bar{x} + 1.65 * s]$
    - 95% CI for X: $[\bar{x} - 1.96 * s, \bar{x} + 1.96 * s]$
    - 99% CI for X: $[\bar{x} - 2.58 * s, \bar{x} + 2.58 * s]$

- Standard Normal Distribution
  - $z = \frac{x-\mu}{\sigma} \sim N(0, 1)$
  - Standardization: observed value-> z-score.
  - Z-score means **how many standard deviation above/below the mean**
  - CDF of standard normal distribution
    - $F(Z) = P(Z < z)$

- the usage of `z-table`, only contains positive z-values
- Transform of negative z-values: $F(-Z) = 1 - F(Z)$

- Shortfall Risk, Safety-first Ratio, Optimal Portfolio Identification
  - Shortfall Risk: $P(R_p < R_L)$
  - Roy's safety-first criterion, for optimal portfolio
    - Function: $minP(R_p < R_L)$
      - R_L: threshold acceptable level return
  - If the portfolio is **normally distributed**, Roy's safety-first criterion
    - Function: $max\ SFRatio$, where $SFRatio = \frac{E(R_p)-R_L}{\sigma_p}$
      - $SFRatio$: the number of std *below* the mean, when calculate the probability, it is the `left-tail` of standard normal distribution in graph. $F(-SFRatio_{max})$
      - SFRatio↑, portfolio return< threshold return↓

3. Log-normal, T, Chi-Square, F-distribution

- Log-normal Distribution
  - Function: e^x, where x is normally distributed.
  - skewed to the right
  - range:$[0, \infty]$, which is used for `asset pricing`
    - price relative: $S1/S0 = 1 + HPR$
      - S_1:end-of-period price
      - S_2:beginning price
      - HPR: holding period return
      - treated as up or down-move(multiplier) terms
  - Calculating **Continuously** Compounded Returns
    - When the compounding periods become shorter and shorter, we use continuously compounding technique
    - Function: $EAR = e^{R_{CC}} - 1$, where R_{CC} is stated annual rate.
    - Financial Calculator: `[value of Rcc]` `[ln]` `[CPT]`
    - $Rcc = ln(1 + HPR) = ln(\frac{S_1}{S_0})$
    - Additive for multiple periods: $HPR_T = e^{Rcc*T} - 1$

- Student T-Distribution
  - bell shape,centered about 0
  - used for small samples(n<30) with approximately normal distribution or large sample(CLT) from populations with unknown variance.
  - Degree of freedom(df): n-1, based on the test of sample means
  - Fatter tails than normal distribution, more area under tails
  - CI for t-dist more `wider` than that of normal dist
  - When n(or df) larger, t-dist more close to standard normal distribution, t-dist become more thinner tails and spiked.
  - more difficult to reject H0 than using normal dist
  - $\alpha_{twoTails} = 2 * \alpha_{oneTail}$
  - critical value increase with df decrease.
- Chi-Square(χ^2)
  - df: $k = n - 1$
  - Asymmetric
  - when df increases, its distribution symmetric increase, approaching normal distribution gradually
  - used for tests whether the values of `variance` are `equal` of normally distributed population
  - range:[-∞,0]
- F-Distribution
  - including two approximately scaled independent chi-square variables, with df m and n
  - $F = \frac{\chi^2/m}{\chi^2/n}$
  - Asymmetric
  - when m and n get larger, the F-distribution become more symmetric and more similar to normal dist.
  - cannot take on negative values
  - usage of F-distribution table
- Monte Carlo Simulation
  - repeated generation of >=1 risk factors with specified parameters(mean,variance,skewness...) to generate a distribution of security values, calculate the mean value as the estimation results.