FRM Notebook 2

计算器
http://www.360doc.com/content/16/0815/12/17753496_583354250.shtml
https://wenku.baidu.com/view/f572b2cf0c22590102029d7b.html

BA II Plus

The Time value of money

- Future value (compounding)
- Present value (discounting)
- Interest rate
  - Discount rate
  - Opportunity cost
- Real risk-free rate 实际无风险利率
  - Has no expectation of inflation in it
  - 就是名义无风险利率扣除通货膨胀影响得到的利率。
- Nominal risk-free rate 名义无风险利率
  - = real risk-free rate + expected inflation rate
  - T-bill (contain inflation premium)
  - (1+R)=(1+r)(1+i)；R--名义利率，r--实际利率，i--通货膨胀率。
- Three types of risk
  - Default risk 拒付风险溢价
    - Cannot pay in a timely manner
  - Liquidity risk 流动性溢价
    - Receive less than fair value if it must be sold for cash quickly
  - Maturity risk 到期风险溢价
    - long-term bonds have more maturity risk
- interest rate = nominal risk-free rate + default risk premium + liquidity premium + maturity risk premium
- single sum cash flow PV
  - Interest rate, discount rate, cost of capital, required rate of return
  - PV = FV / (1+I/Y)^N
- Annuities 年金
  - Equal cash flow, equal intervals
  - Ordinary annuity: at the end of each compounding period
- Perpetuity
  - a fixed amount of money at set intervals over an infinite period of time.
    - British consul bonds, preferred stocks (dividend payment)
    - PV = PMV / (I/Y)
- PV and FV of uneven cash flow
- Compounding period other than annual
  - Increase effective rate of interest, increase FV, decrease PV
  - **effective** rate of interest = (1+r/m)**(m*y) – 1

Probabilities

- probability function f(x)
    - discrete distribution
    - continuous distribution p(x)=0, range P(x1<=x<=x2)>0
- probability density function (pdf)
- cumulative distribution function (cdf)
- inverse cumulative distribution function
- conditional probabilities
- joint probability P(AB)= P(B) P(A|B)
- independent events
    - P(AB) = P(A) P(B)
    - P(A|B) = P(A)
- Mutual exclusive events
    - P(AB) = 0
- Add
    - P(A or B) = P(A) + P(B) – P(AB)
- Independent events
    - P(X_1 or X_2 …) = \sum P(X_i)
    - P(X_1 and X_2 …) = \prod P(X_i)
- Probability Matrix
    - Marginal probability

Statics

- Mean (population or sample)
- Median: the middle point of a data set when data is arranged in ascending order.
- Mode: the value that occurs most frequently
- Geometric mean: G = (X1 X2)^(1/n)
    - Compound annual rate
    - 1+R_G = ((1+R_1) (1+R_2) … (1+R_n))^(1/n)
- expected value
    - sum P_i * X_i
    - E(XY) = E(X) E(Y) independent events
- Variance
    - Var(X) = E((X-u)^2) = E(X^2) – E(X)^2
    - Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) +2 abCov(X,y)
    - Var(X+Y)= Var(X-Y) = Var(X)+Var(Y) independent events
- Covariance and Correlation
    - Cov(X,Y) =E((X-U_x)(Y – U_y)) = E(XY) – E(X)E(Y)
    - COV(X,Y) = 0 for independent events
- Correlation
    - Corr(X,Y) = Cov(X,y) / (sig(X) sig(Y))
    - Linear relation -1, 1
- Moments and Central Moments
    - Raw moments: E(X^k) = \sum p* x^k
        - U = E(X)
    - Central moments: E((X-U)^k) = \sum p * (x - u)^k

        - Var(X) = E((X-u)^2)
    - Skewness
        - Skewness= E((X-u)^3) / sig^3
        - 0: symmetric
        - positive: right tail
            - mode < median < mean
        - negative: left tail
            - mean < median < mode
    - Kurtosis
        - Kurtosis = E((X-u)^4) / sig^4
            - Leptokurtic: >3, more peaked and more tail than normal
                - A greater percentage of small deviation from mean
                - A greater percentage of extremely large deviation from mean
            - platykurtic: < 3, less peaked and less tailed than normal
            - mesokurtic: =3
    - Excess kurtosis
        - Excess kurtosis = kurtosis – 3
- Best linear unbiased estimator (BLUE)
    - Unbiased
        - Expected value of sample mean = population mean
    - Efficient
        - Variance of sampling distribution is smallest
    - Consistent
        - Accuracy increase as the sample size increases
    - Linear
        - A linear combination of sample data

Distributions

- Parametric and nonparametric
- Uniform
    - f(x) = 1 / (b-a)
    - F(x) = (x - a) / (b - a)
    - P(x1<=x<=x2) = (x2-x1)/(b-a)
    - E(x) = (a+b) / 2
    - Var(x) = (b-a)^2 / 12
- Bernoulli
    - Success 1 with p, failure 0 with 1-p
    - E(x) = p
    - Var(x) = p(1-p) = pq
- Binomial
    - P(x) = (n choose x) p^x (1-p)^(n-x)
        - n! / ((n-x)! x!)
    - E(x) = np
    - Var(x) = npq
- Poisson
    - X: the number of success per unit
    - Lambda: the average number of success per unit

  - P(X=x) = lamba^x e^(-lambda) / x!
  - Mean=variance = lambda
- Normal
  - f(x) = exp(- 0.5* (x-u/sigma)^2) / sqrt(2 pi sigma^2)
  - N(u, sigma^2)
  - Skewness = 0, mean=median=mode
  - Kurtosis = 3
  - Linear combination is also normally distributed
  - Confidence interval
    - 68% within one std
    - ~95% within two std
    - 90% within 1.65
    - **95% within 1.96**
    - 99% within 2.58
  - Standard Normal Distribution
    - z = (x – u) / sigma
  - z-table
    - F(-Z) =1- F(Z)
- Lognormal
  - exp^x, where x is normally distributed
  - skew to right, positive
  - lognormal: price relative P1/P0 = (1+r)
- Central Limit Theory
  - Population: u, sig^2
  - Sampling distribution (n samples each time)
    - U, sig^2/n
    - Sufficient large: n >= 30
  - If n is sufficiently large (n>=30), the sampling distribution of sample means will be approximately normal.
- Student's t-Distribution 样本均值（数量小，或者方差未知）
  - Arise from sampling
    - Sample mean X' = sum X_i / n
    - Sample variance S^2 = sum (X_i - u)^2 / (n-1)
    - X' – u / (sig/sqrt(n)) -> normal distribution
    - (X' - u) / (s/sqrt(n)) –> n-1 t-distribution
  - 数据量少，方差不知道。
  - Used to construct confidence intervals based on small samples (n<30)
  - For population with **unknown variance** and a normal or approximately normal distribution
  - Symmetrical, one parameter (degrees of freedom, n-1)
  - More probability in the tails than normal
  - n gets larger, the shape approaches a normal distribution
- Chi-Squared distribution
  - 定义：平方和， Q= sum e_i^2 , where e_i 是标准正态分布
  - X^2_(n-1) = (n-1) s^2 / sig_0^2
    - s^2: sample variance
    - sig_0^2: hypothesized value for the population variance

  - right skew, positive, n gets large, approach a normal distribution
- F-distribution
  - Ratio of sample variance
  - $F=s1^2/s2^2$ $(n1-1,n2-1)$
- Mixtures
  - Skewness: different mean
  - Kurtosis: different variances

Bayesian Analysis
- $P(A|B) = P(A) P(B|A) / P(B)$
  - Prior probability $P(A)$, posterior probability $P(A|B)$
- Frequentist
  - questionable with a small sample size
- Bayesian
  - a beginning assumption regarding probabilities

Hypothesis Testing and Confidence Intervals
- sample error
  - difference between sample statistic and population parameter
- sample statistic – random variable
- sampling distribution – distribution
- population variance $sigma^2 = … / n$
- sample variance $s^2= … / (n-1)$
- standard error (std of the sample means)
  - $sig_x = sig / sqrt(n)$, population
  - $s_x = s / sqrt(n)$, sample
- Confidence Interval
  - Level of significance: alpha
  - Degree of confidence: 1-alpha
  - Point estimate +- reliability factor * standard error
- Normal distribution with known variance (Z)
  - $x + z_{(a/2)} sig/sqrt(n)$
  - 1.65, for 90% confidence
  - 1.96 for 95%
  - 2.58 for 99%
- Normal with unknown variance (T)
  - $X + t_{(a/2)} s / sqrt(n)$
- NonNormal with
  - known variance: Z- distribution
  - unknown variance: T- distribution
  - $N >= 30$ because of CLT
- Hypothesis testing
  - Null and alternative hypothesis about **population**.
  - Test statistic
  - Critical value (level of significance)

- Test statistic = (sample statistic – hypothesized value) / stand error of sample statistic
- One-tailed test and two-tailed test
  - Two tailed mean test
  - Decision rule: test statistic > upper critical value or < lower critical value
- Critical value
  - 5% significance level
    - two-tailed: 1.96
    - one-tailed: 1.645
- Type error
  - Type 1: rejection of null when it is actually true
    - Significance level
  - Type 2: failure to reject null when it is actually false
    - 1- Power of a test
- Power of a test
  - Probability of correctly rejecting the null when it is false
- Reduce significance level from 5% to 1%, increase type II error and decufese power of test
- For a given significance level, increase sample size, decrease type II error

- P-value = P(as extremely as test statistic |H_0 is true)
  - P(x > test statistic) and/or P(x < -test statistic)
  - Minimal significance level to reject the null hypothesis
- T-test
  - Variance is unknown and either of the following condition exist
    - Sample size is large (>=30)
    - Sample is small (<30), but it is normally or approximately distributed
- Z-test
  - Normally distributed with known variance
  - 10% 1.65, 1.28
  - 5% 1.96, 1.65
  - 1% 2.58, 2.33
  - if sample size is large, unknown variance, Z-test can also be used
- Chi-Squared test
  - Variance of a normal distribution
  - Statistic = (n-1)s^2 / sigma^2
- F-test
  - variances of two populations
  - normal and I.i.d.
  - F= s1^2 / s2^2
  - Degree of freedom n1-1, n2-1
  - 总是把大的放在分子上，意味着只需要考虑 right-side tail 的。
- Chebyshev Inequality
  - P(|x-u|< k *s) >= 1- 1/k^2, any distribution
- Backtesting
  - VaR usually 95% confidence level
  - Exception are serially correlated

- Correlated with overall market volatility
  - Fail to react quickly to changes in risk levels

Linear Regression with One Regressor

- Coefficient, Intercept, slope
- Population regression function
  - Error term: epsilon = Y – E(Y|X), nonsystematic or random component
- sample regression function
  - residual: e = Y – f(.)
- OLS
  - Sum of residuals \sum e^2
  - B1 = Cov(X,Y) / Var(X)
  - B0 = E(Y) – B1 E(X)
- Assumption
  - E(e|X)=0, Var(e|X) = constant, normally distributed
  - All (X,Y) are iid
  - No outlier
- Properties
  - Unbiased, consistent, and efficient under special conditions
  - Sample, 100 to get consistent
- Results
  - SSR/SSE: sum of squared residuals
  - R^2: coefficient of determination
  - TTS/SST: total sum of squares around the mean
  - ESS/RSS: Explained sum of squares
  - TSS = ESS + SSR
  - Sum (y – E(y))^2 = SSR sum (y-f(.))^2 + ESS sum(f(.)-E(y))^2
- R^2
  - R^2 = ESS / TSS = 1 – SSR/TSS
  - Correlation coefficient r = sqrt(R^2) for two-variable regression
- Standard error of regression (SER)

Hypothesis Test

- Regression coefficient
  - Confidence interval
    - u + t * s
    - Degree n-2
  - Hypothesis Test
    - (b1 – B) / s
    - significance level -> critical value
  - p-value
    - the minimal significance level to reject null hypothesis
- Predicted Values
  - Y= a + b*x
  - df = n -2
- homoskedasticity and heteroscedasticity 残差
  - Homoscedasticity

        - Variance of residual is constant
    - heteroscedasticity
        - unconditional:
        - conditional: depend on x
        - affects
            - standard errors are unreliable
            - coefficient estimates aren't affected
- The Gauss-Markov theorem
    - if the linear regression model assumptions are true and the regression errors display homoskedasticity,
- Small Sample Size
-

Linear Regression with Multiple Repressors

- omitted variable bias
- intercept term and partial slope coefficients
- homoscedasticity (variance of error term is constant for all variables) and heteroscedasticity
- measure of fit
    - standard error of the regression (SER)
        - SER^2 = SSR/(n-k-1), where SSR is sum of squared errors \sum_i (Y_i – Y_i^)^2
    - Coefficient of Determination R^2
        - R^2 = ESS / TSS = 1 – SSR / TSS
            - TSS = sum_i (Y_i – E(Y))^2
    - Adjusted R^2
        - R_a^2 = 1 – (n-1)/(n-k-1) * (1-R^2)
            - -k / (n-k-1) + (n-1)/(n-k-1)*R^2
- Assumption
- Multicollinearity
    - Two or more variables, or their linear combination, are highly correlated with each other.
    - Distorts standard errors and coefficient standard errors
- Detect Multicollinearity
    - T-test shows none of coefficients is significantly different than 0, while the R^2 is high.
    - Exact two: High correlation among variables. >0.7
- Correcting Multicollinearity
    - Omit one or more
    - Stepwise regression

Hypothesis Tests

- Regression coefficient – t-statistic
    - T = (b_i – B_i) / s_i, df=n-k-1
- P-value
- Confidence interval

- Joint Hypothesis Test – F-statistic
    - All coefficients are equal to 0.
    - Test whether at least one variable is significant
    - One-tailed test, df=(k, n-k-1)
        - F = (ESS/k) / (SSR / (n-k-1))
        - ESS = TSS - SSR
- R^2
    - R2 = 1- SSR/TSS
- Specification Bias
- Test of a single restriction involving multiple coefficients
    - Software
    - Single null hypothesis

Modeling and Forecasting Trend

- Linear and nonlinear
    - Quadratic
    - Exponential trend: log-linear
- Model Selection Criteria
    - MSE = SSR / N
    - R^2 = 1 – MSE*N / TSS
    - Model selection: out-of-sample
    - s^2 = SSR / (N - k)
    - adjusted R^2 = 1 – [SSR / (N-k)] / [TSS / (N-1)]
    - Akaike and Schwarz Criterion
        - S = T / (T-K) * MSE
    - AIC
        - AIC = exp(2k/T) * MSE
    - SIC
        - SIC= T^(k/T) MSE
- Consistency
    - SIC most consistent
    - AIC asymptotic efficiency
        - variance close to true model

Modeling and Forecasting Seasonal

- Seasonality in a time series is a pattern that tends to repeat from year to year.
- Two ways
    - using a seasonally adjusted time series
    - regression analysis with seasonal dummy variables.
- Seasonal dummy variable
    - Intercept and s-1 dummy variables to avoid multicollinarity
    - Holiday variations (HDV) and trading-day variations (TDV)

Characterizing Cycles

- Covariance stationary process 过去未来关系稳定
    - **Mean** must be stable over time
    - **Variance** must be finite and stable over time

  - **Covariance** structure must be stable over time
    - Tau, lag or displacements
- Autocovariance
  - Covariance between current and t periods in the past
    - Stable means covariance depends on tau not time
- Autocorrelation function
  - Autocovariance / variance
- Autoregression
  - Linear regression of a time series against its own past values
  - **Regression** coefficient is partial auocorrelation at that lag
  - Partial autocorrelation function
- Not covariance stationary
  - Cannot model it directly
  - Solution: identify and isolate an underlying covariance stationary process
    - Remove: Such as trend, seasonality.
    - Transformation: difference, logarithmic scaling
- White noise
  - Serially uncorrelated: No correlation among any of its lagged values
  - Zero-mean white noise: mean =0, variance is constant
  - Independent white noise: observations are independent and uncorrelated
    - Next value has no conditional relationship to any of its past values
  - Gaussian/normal white noise: follow a normal distribution
  - Model errors should be a white noise process
- Lag operators 利用条件依赖来预测
  - L one period before, y_(t-1)=L t_t
  - L^m y^t = y_(t-m)
  - Distributed log: different weights
  - Lag operator polynomial of degree 3.
- World's theorem
  - A covariance stationary process can be modeled as an infinite distributed lag of a white noise process 无限白噪声之和
    - e_t + b_1 e_(t-1)+ … = \sum_i b_i e_(t-i)
    - generally linear process
    - e: innovations, errors, not necessary independent.
  - If innovations have a conditional relationship with past innovations
  - Can be approximated with a ratio of relational distributed lags
- Estimating Autocorrelations
  - autocorrelation(tao) = \sum *{t=tao+1} (*y_t_u)(y_(t-tao) - u) / \sum (y_t-u)^2
  - correlogram or sample autocorrelation function.
  - If T observations, the std is 1/sqrt(T),
    - Within the band +-2/sqrt(T), then white noise with 95% confidence
  - White noise Hypothesis: autorrelations are jointly equal to 0.
  - Box-Pierce Q-statics: chi-squared distribution
  - Ljung-Hox Q-statics: similar but useful with small samples

Modeling Cycles: MA, AR, and ARMA Models

- First-order Moving average

- **Moving average repr**: y_t = e_t(random shock) + \theta e_(t-1) (lagged unobservable shock)
- autocorrelation cutoff 相关截止, 自相关函数是截尾的
  - p = theta / (1+theta^2), where for p=0 for tao >1
- autoregression repr: e_t = y_t – theta e_(t-1)

- MA(q): **random shocks 未观察值**
  - y_t = e_t + theta_1 * e_(t-1) + … theta_q * e_(t-q)
- first-order AR
  - must have a zero mean and a constant variance
  - y_t = theta y_(t-1) + e_t
    - |theta| <1
  - Yue-Walker equation 相关衰, 偏相关函数是拖尾的
    - Correlation_t = theta^t, for t=0,1,2,
  - If theta<0, decay but oscillate
- AR(p) 观察值，可以捕捉季节
  - y_t = phi_1 y_(t-1) +… + phi_p y_(t-p) + e_t
  - it is covariance stationary if |phi| < 1
- Autoregressive moving average process (ARMA(p,q))
  - Stock price unobserved shocks (MA part) and their own lagged behavior (AR)
  - y_t= \phi y_(t-1) + e_t + theta e_(t-1)
- example
  - AR(2)
  - seasonally effect ARMA(3,1)

Volatility

- volatility 连续复利，波动性
  - std of its continuously compounded return
  - option: std of return over a one-year period
  - risk: std of daily return
  - **continuously compounded return E = S exp(r)**
    - u_i = ln (S_i / S_(i-1)) , where S is the asset price
  - Proportional change E = S (1+r), r = (E_S)/S
    - U = (S_i – S_(i-1)) / S_(i-1)
  - Daily volatility s, then over N days is sqrt(N)*sigma
  - 252 days (business days), sqrt(252)=15.87
- variance rate: **square of volatility**
  - sigma^2, N days is N*sigma^2
- implied volatility 隐含波动率
  - option, option pricing model, BSM
  - model price = market price
  - volatility index (VIX), fear index
- power law
  - change in price is normal distribution
  - P(V>X) = K X^(-alpha)
- Estimation volatility
  - Mean of u, mean = \sum u_i / m
  - Variance: sig^2 = \sum u_i^2 / m

- Weighted
  - Sig^2 = \sum alpha_i * u_i ^2 / m
- autoregressive conditional heteroskedasticity model, ARCH(m),
  - Sig^2 = r * V_l + \sum alpha_i * u_i ^2 / m
  - R + sum alpha_i=1
  - V_l: Long-run variance

- exponentially weighted moving average (EWMA)
  - decline exponentially back through time
  - **sig_n^2 = lambda * sig_(n-1)^2 + (1-lambda) u_(n-1)^2**
- Generalized autoregressive
  - GARCH(p,q)
    - P Number of lagged terms on historical **returns squared**
    - Q: number of lagged terms on historical **volatility**.
  - GRACH(1,1)
  - most recent estimates of variance and squared return, but also a variable that accounts for a long-run average level of variance.
  - **Sig_n^2 = gamma * V_L + alpha * u_(n-1)^2 + beta sig_(n-1)^2**
    - W = gamma * V_l
    - Gamma = 1- alpha – beta
  - **V_L = w / (1-alpha -beta)**
  - **Stable: Alpha + beta < 1**
  - EWMA is a special case, use it when instability
  - variance tends to revert to a long-term average level.
  - mean-reverting: option
  - use MLE to estimate, guess and test
- Mean Reversion
  - **Persistence**: alpha + beta
    - Higher: take longer time to revert
    - If persistence =1, no mean reversion
    - If persistence >1, use EWMA

Correlations and Copulas

- Correlation
  - COV(X,Y)= E(XY) – E(X) E(Y)
- EWMA
  - Cov = lambda cov + (1-lambda)X*Y
  - Var = lambda var + (1-lambda) X^2
- GRACH(1,1)
  - Cov = w + alpha * X*Y+ beta cov
- Consistency
  - Variance-Covaraince matrix
  - Positive-semidefinite is internally consistent
    - x^T W x >= 0, for all x
  - Variance and covariance rates must be calculated using the **same** EWMA or GARCH model and **parameters** to ensure that a positive-semidefinite model is constructed. For example, if an EWMA model is used with X = 0.90 for

estimating variances, the same EWMA model and X should be used to estimate covariance rates.

- Generate samples
    - E_y = r * Z_x + sqrt(1-r^2)*Z_y
    - The expected value of Yis therefore linearly dependent on the conditional value of X.
- Factor models
    - U_i = alpha_i * F + sqrt(1-alpha^2) * Z_i
        - N(0,1)
        - -1 <= Alpha_i <=1, 1-alpha^2>=0 =>
        - F ~N(0,1) and Z_i ~N(0,1)
        - Z_i are independent
        - F and Z_i are independent
    - Pros
        - positive-semidefinite
        - N estimations for correlation instead of N(N-1)/2
        - CAPM
- Copulas
    - If both X and Y's marginal distributions are normal, then we can assume the joint distribution of the variables is bivariate normal. However, if the marginal distributions are not normal, then a **copula** is necessary to define the correlation between these two variables.
    - A copula creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions
    - **Gaussian copula** maps the marginal distribution of each variable to the standard normal distribution N(0,1). Based on percentile.
    - preservation of the original marginal distributions while defining a correlation between them.
- Types
    - Gaussian copula
    - Student's t-copula
    - Multivariate copula
    - Factor copula
        - U_i have student's t-distribution if Z_i is normal and F is student's t
        - Used to calculate VaR
- Tail Dependence
    - There is a greater tail dependence in bivariate student's t-distribution than bivariate normal distribution

Simulation Methods

- 参考
    - http://www.docin.com/p-102193234.html
- Monte Carlo
    - model complex problems or to estimate variables when there are small sample sizes.

- pricing exotic options, estimating the impact to financial markets of changes in macroeconomic variables, and examining capital requirements under stress-test scenarios.
- Steps
  - Specify the data generating process (DGP)
    - Probability distribution
  - Estimate an unknown variable or parameter
  - Save the estimates from step 2
  - Go back to step 1 and repeat this process N times

- Reducing Sampling Error
  - Standard error: s/sqrt(N)
  - Increase the number: four times reduce by half
  - Reduce standard error
- Antithetic variate
  - Theory
    - X = (x_1 + x_2)/2,
    - Variance = (var(x1)+var(x2)+2cov(x1,x2))/4
    - • If x_1 and x_2 are **negative** correlated, variance is smallest
  - method
    - sample u, complement –u
- Control variate
  - A control variate involves replacing a variable x (under simulation) that has unknown properties with a **similar** variable y that has **known** properties.
    - x with unknown properties and
    - y control variable with known properties
    - x* = y + (x' – y'), where x' and y' are Monte Carlo estimates
    - x* has smaller variance than x
    - have similar properties to y
  - var(x*) = var(y) + var(x'-y') + 2cov(y, x'-y') # y is known
    - = var(x'-y')
    - =var(x') + var(y') – 2cov(x',y')
  - requires cov(x',y')>var(y')/2, then var(x*) < var(x')
  - corr(x',y') > 0.5 sqrt(var(y') / var(x'))
- Reuse Random Number
  - Dickey-Fuller test (used to determine whether a time series is covariance stationary)
    - the sampling variability is reduced, but the accuracy of the actual estimates is not increased.
  - for different experiments with options using time series data.
    - Test difference among options (maturity)
- Bootstrapping
  - Use historical data (draw with return)
  - data sets with approximately the same distribution properties as the original data
  - any dependency of variables or autocorrelations in the original data set will no longer be present
  - inefficient

        - outlier
            - repeat many times
        - non-independent data
            - moving block bootstrap
- Random number generation
    - $y_{i+1} = (a\ y_i + c)$ modulo m, i=0,1,2,..,T
    - y_0: seed, influence the early numbers
- Disadvantages
    - High computation cost
    - Results are imprecise
        - Assumption of model inputs or the probability distribution
    - Results are difficult to replicate
        - No seed, use more replications
    - Results are experiment-specific