1 Inventory Management

1.1 Inventory System

Inventory is a physical resource that a firm holds in stock with the intent of selling it or transforming it into a more valuable state.

Inventory system is a set of policies and controls that monitor levels of inventory and determine what levels should be maintained, when stock should be replenished, and how large orders should be.

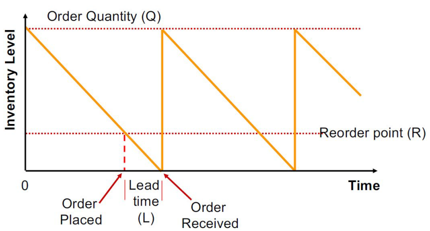
Decisions:

1. *Order quantity* (Q): How much should we order?
2. *Reorder point* (ROP; R): When should we place an order?
3. *Safety stock* (SS): How much safety stock should we maintain?

Objectives:

1. Maximize the level of *customer service* by avoiding understocking.
2. Minimize the *cost* of providing an adequate level of customer service.
3. Maximize the *profit*.

The Inventory Cycle:



1.2 Inventory Management Under Deterministic Demand

Batch Sizing: Determination of Q when ordering.

1. Using a large order size (i.e., ordering infrequently): we suffer a large inventory *holding cost*.
2. Using a small order size (i.e., ordering frequently): we suffer a large *fixed cost of ordering*.

The order quantity that minimizes the total cost per period is called the *Economic Order Quantity* (EOQ).

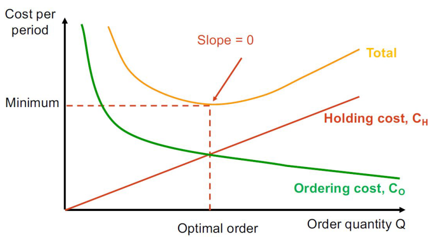
Total Cost Formula:

1. The holding cost depends on average stock: $Q/2$.
2. Ordering cost depends on number of orders per period: D/Q.

The total cost per period formula as a function of the batch size ordered (Q):

$$T(Q) = \frac{Q}{2}C_H + \frac{D}{Q}C_O$$

D is the demand per period, Q is the batch (lot) size, C_O is the (fixed) cost of placing one order and C_H is the cost of holding one item in store for one period (financing and physical storage).



Find the Q^* that minimizes $T(Q)$:

$$EOQ = Q^* = \sqrt{2D\frac{C_O}{C_H}}$$

Reorder Point: The point in time by which stock must be ordered to replenish inventory before a stockout occurs.

$$R = dL + SS$$

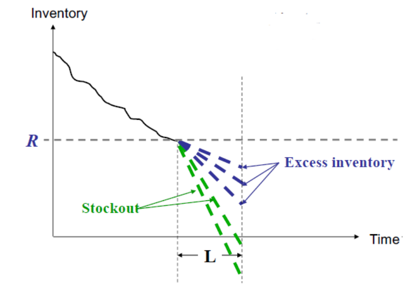
R is the reorder point, d is the average demand per time period (constant), L is the number of time periods between placing order and delivery and SS is the safety stock.

1.3 Inventory Management Under Stochastic Demand

(Q, R) Decisions:

1. We choose R to meet the demand during *lead time*.
2. Tradeoff in Q: Fixed cost versus holding cost.

Demand is *random* and *stationary*. Expected demand is λ per unit time.



We assume that D represents the *demand during the lead time* and has probability distribution $f(x)$ with mean μ and standard deviation σ. Average inventory level before an order arrives is $s = R - \mu$.

$$T = \frac{Q}{\lambda}$$

Expected shortage per cycle is:

$$n(R) = \int_0^R 0 f(x)dx + \int_R^\infty (x - R)f(x)dx$$
$$= \int_R^\infty (x-R)f(x)dx = \sigma L\left(\frac{R-\mu}{\sigma}\right)$$

Standard loss function $L(z) = \phi(z) - z(1 - \Phi(z))$.

(Q, R) Model: The expected total cost per unit time = fixed cost + holding cost + stockout (backorder) cost.

$$C(Q) = h\left(s + \frac{Q}{2}\right) + \frac{K}{T} + p\frac{n(R)}{T}$$
$$= h\left(\frac{Q}{2} + R - \lambda L\right) + K\frac{\lambda}{Q} + p\frac{\lambda n(R)}{Q}$$

K is the setup cost per order, h is the holding cost per unit per unit time, c is the purchase price (cost) per unit and p is the penalty cost per unit of unsatisfied demand.

Optimal solution:

$$Q = \sqrt{\frac{2\lambda[K + pn(R)]}{h}}$$
$$F(R) = 1 - \frac{Qh}{p\lambda}$$

1.4 Newsvendor Model

Newspapers are *perishable* and daily sales are represented by the random variable D.

Newsvendor Model: The expected cost = total overage (not enough demand) + underage cost (too much demand):

$$G(Q, D) =$$
$$c_o \max(0, Q - D) + c_u \max(0, D - Q)$$

c_o is the unit cost of overage and c_u is the unit cost of underage.

The expected cost function $G(Q) = \mathbb{E}[G(Q, D)]$ is given by:

$$\int_0^Q c_o(Q - x)f(x)dx + \int_Q^\infty c_u(x - Q)f(x)dx$$

The optimal order quantity Q^* satisfy:

$$\boldsymbol{F(Q^*) = \frac{c_u}{c_u + c_o}} \quad [= \mathbb{P}(D \le Q^*)]$$

2 Decision Analysis

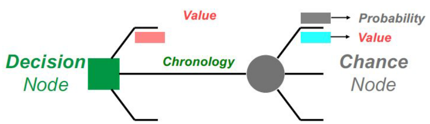
2.1 Decision-making Framework

Different decisions rules lead to different optimal actions based on *payoff tables*:

1. *Maximax:* Maximize the maximum payoff (i.e., under the best-case scenario). High risk, high return strategy.
2. *Maximin:* Maximize the minimum payoff (i.e., under the worst-case scenario). Low risk, low return strategy.
3. Maximize *Expected Monetary Value* (EMV) using the probabilities for each *state* of nature.

2.2 Decision Trees

Structure problem as a sequence of decisions and outcomes.



Best decision: Roll the payoffs back from the leaves to the root of the tree based on decision rules such as:

1. To overcome an event node, find *average* payoff over all possible events.
2. To overcome a decision node, take decision with *maximum* payoff.

Results in action plan for each future contingency with *expected monetary value* or payoff (EMV).

2.3 Sensitivity Analysis

If the values for purchases, profits and probabilities are not accurate, we can assume that the probability is p. The optimal decision then depends on the probability p.

2.4 Expected Value of Perfect Information

Expected value of perfect information (EVPI): The maximum amount willing to pay to know the future.

$$\text{EVPI} = \text{EMV w/ PI} - \text{EMV}$$

Expected value of sample/imperfect information (EVSI): The maximum amount willing to pay to get an expert's opinion given his/her ability to understand the environment, denoted by q.

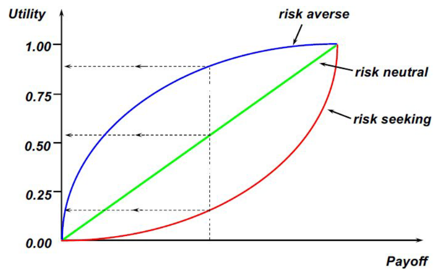
Let a be the event that the expert says A and b the event that the expert says B.

Bayes' Rule: Given $P(a \mid A) = q_1$ and $P(b \mid B) = q_2$:

$$P(A \mid a) = \frac{P(a \mid A)P(A)}{P(a)}$$
$$= \frac{P(a \mid A)P(A)}{P(a \mid A)P(A) + P(a \mid B)P(B)}$$

2.5 Risk Attitude

Risk attitude can be captured by *utility functions*.



A *risk averse* decision maker assigns the largest relative utility to any payoff but has a diminishing marginal utility for increased payoffs.

A *risk seeking* decision maker assigns the smallest utility to any payoff but has an increasing marginal utility for increased payoffs.

A *risk neutral* decision maker falls in between these two extremes and has a constant marginal utility for increased payoffs.

3 Dynamic Programming

3.1 Stagecoach Problem

The *greedy algorithm* is fast with complexity of $O(NK)$, where N and K are the number of stages and number of nodes in each stage, respectively. However, it does not guarantee optimality.

An *exhaustive search* can guarantee an optimal solution, but it is time consuming with complexity of $O\left(K^N\right)$.

The *dynamic programming* approach is based on backward inductions starting at the last stage and it guarantees an optimal solution with complexity of $O\left(NK^2\right)$:

1. Start from the last stage and work backward.
2. At each node of a particular stage, find an optimal route to location J based on the optimal routes to location J from the child nodes of the current node.
3. The algorithm terminates when the first stage is done.

3.2 Sequential Decision Models

System dynamics:

- n: stage index
- s_n: state of the system in stage n
- $s_{n+1} = g_n(s_n, x_n)$: state transition (how the system state moves from stage n to stage $n+1$ if action x_n is taken in stage n)
- $c_n(s_n, x_n)$: system cost in stage n
- $c_N(s_N)$: system cost in stage N

The system optimization problem is the following minimization problem:

$$\min_{x_n} c_N(s_N) + \sum_{n=0}^{N-1} c_n(s_n, x_n)$$

Define new reward-to-go functions:

$$f_n(s_n, x_n) = c_n(s_n, x_n) + f_{n+1}^*(g_n(s_n, x_n))$$
$$f_n^*(s_n) = \min_{x_n} f_n(s_n, x_n)$$

We obtain an *optimality condition* (Bellman equation):

$$f_n^*(s_n) = \min_{x_n} c_n(s_n, x_n) + f_{n+1}^*(g_n(s_n, x_n))$$

3.3 Deterministic Dynamic Programming

Dynamic Programming is based on the *principle of optimality*: A truncated optimal solution is also an optimal solution to the truncated sub-problem.

In deterministic dynamic programming, a pair of state and action in stage n leads to a unique state in stage $n+1$.

The optimality conditions are:

$$f_n^*(s_n) = \min_{x_n \in U_n} c_n(s_n, x_n) + f_{n+1}^*(g_n(s_n, x_n))$$

The constraints in stage n are defined by U_n.

3.4 Probabilistic Dynamic Programming

In probabilistic dynamic programming, the state in stage $n+1$ is governed by a probability distribution $p_{nt}(s_n, x_n)$ that is completely determined by state s and action x in stage n.

In the optimality equations, we solve an optimization problem by minimizing the expected value:

$$f_n^*(s_n) = \min_{x_n} \sum_{t=1}^{S} p_{nt}(s_n, x_n)\left(c_n(s_n, x_n, t) + f_{n+1}^*(t)\right)$$

When $S = 1$, we recover deterministic dynamic programming.

4 Markov Chains

4.1 Stochastic Processes

Stochastic process is a collection of indexed random variables (RVs) $X_t, t \in T$ which are typically statistically dependent.

State space is the set of possible values of X_t's.

4.2 Discrete Time Discrete State Stochastic Processes

Discrete space: X_t's are discrete RVs $\{X_1, X_2, X_3, \ldots\}$.

Discrete time: $T = \{0, 1, 2, 3, \ldots\}$.

Markovian property: The probabilities that govern a transition from state i at time t to state j at time $t+1$ only depend on the state i at time t and not on the states the process was in before time t.

$$\begin{aligned} \mathrm{P}(X_{t+1} = j \mid X_0 = i_0, \ldots, X_t = i_t) \\ = \mathrm{P}(X_{t+1} = j \mid X_t = i_t) \end{aligned}$$

Transition probabilities are called *stationary* if:

$$\mathrm{P}(X_{t+1} = j \mid X_t = i) = \mathrm{P}(X_1 = j \mid X_0 = i)$$

The stationary transition probabilities are conveniently stored in a *transition matrix*.

$$P_{ij} = \mathrm{P}(X_1 = j \mid X_0 = i)$$

4.3 Markov Chains

MC is completely characterised by transition probabilities P_{ij} from state i to state j that are stored in an $n \times n$ transition matrix P.

Rows of transition matrix sum up to 1. Such a matrix is called a *stochastic matrix*.

Initial distribution of states is given by an initial probability vector:

$$q^{(0)} = \left(q_1^{(0)}, \ldots, q_n^{(0)}\right)$$

Chapman-Kolmogorov Equations: The n-step transition probabilities $P_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$ obey the following law (for arbitrary $0 < m < n$):

$$P_{ij}^{(n)} = \sum_k P_{kj}^{(n-m)} P_{ik}^{(m)}$$

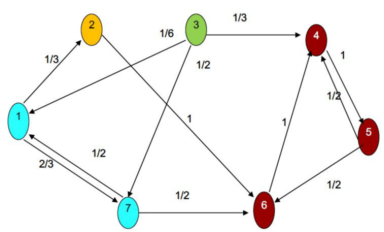
The n-step transition probability matrix $P^{(n)}$ is the n-th power of the 1-step transition probability matrix P:

$$P^{(n)} = P^n$$

4.4 The Transition Network

A stochastic matrix is said to be *irreducible* if each state is accessible from each other state.

Classes of States: State i and j communicate if i is accessible from j and j is accessible from i. Communicating states form *classes*. The state space may be partitioned into disjoint classes.



A class is called *absorbing* if it is not possible to escape from it.

A class A is said to be *accessible* from a class B if each state in A is accessible from each state in B.

4.5 First Passage Times

The first passage time from state i to state j H_{ij} is the number of transitions until the process hits state j if it starts at state i.

Define $f_{ij}(k) = \mathrm{P}\left(H_{ij} = k\right)$, probability that the first passage from state i to state j occurs after k transitions.

The expected first passage time is

$$\mathbb{E}\left(H_{ij}\right) = f_{ij}(1) + 2f_{ij}(2) + 3f_{ij}(3) + \ldots$$

The law of total probability gives:

$$\begin{aligned} P_{ij}^{(n)} = P_{jj}^{(n-1)} f_{ij}(1) + \cdots \\ + P_{jj} f_{ij}(n-1) + f_{ij}(n) \end{aligned}$$

Hence $P_{ij}^{(n)} = \sum_k P_{jj}^{(n-k)} f_{ij}(k)$ results in the recursive formula:

$$f_{ij}(1) = P_{ij}$$
$$f_{ij}(2) = P_{ij}^{(2)} - f_{ij}(1) P_{jj}$$
$$\vdots$$

The law of total probability shows:

$$\begin{aligned} \mathbb{E}\left(H_{ij}\right) &= \sum_k \mathbb{E}\left(H_{ij} \mid C_k\right) P_{ik} \\ &= P_{ij} + \sum_{k \neq j}\left(1 + \mathbb{E}\left(H_{kj}\right)\right) P_{ik} \\ &= 1 + \sum_{k \neq j} \mathbb{E}\left(H_{kj}\right) P_{ik} \end{aligned}$$

Hence, we can find expected first passage times by solving systems of equations.

4.6 Long-Term Behaviour

Distribution of X_n: The probability vector for X_n is:

$$q^{(n)} = q^{(n-1)} P = \ldots = q^{(0)} P^{(n)}$$

As n tends to infinity:

$$\lim_{n\to\infty} q^{(n)} = \lim_{n\to\infty} q^{(0)} P^{(n)} = q^{(0)} \lim_{n\to\infty} P^{(n)}$$

The limit may not exist when the process has periodic behaviour. *Period* of a state i is equal to the greatest common divisor of n such that $P_{ii}^{(n)} > 0$

Aperiodicity: A state with period 1 is called *aperiodic*. State i is aperiodic if and only if there exists N such that $P_{ii}(N) > 0$ and $P_{ii}^{(N+1)} > 0$.

4.7 Steady-State Distributions

Aperiodicity is a class property, i.e., if one state in a class is aperiodic, then so are all others.

If a Markov chain is irreducible and aperiodic, then all limits exist:

$$\overline{P}_{ij} = \lim_{n\to\infty} P_{ij}^{(n)}$$

All rows of its long-term transition probability matrix $\overline{P}$ are identical to the unique solution $\pi = (\pi_1, \ldots, \pi_m)$ of the equations:

$$\pi P = \pi, \quad \sum_{i=1}^{m} \pi_i = 1$$

The probability vector π is called the *steady-state* (or *stationary*) probability distribution of the Markov chain.

If process hits state i, a payoff of $g(i)$ is realized, then the average payoff per period after n transitions is:

$$Y_n = \frac{g(X_1) + \ldots + g(X_n)}{n}$$

Long-run expected average payoff per time period:

$$\lim_{n\to\infty} \mathbb{E}(Y_n) = \sum_{j=1}^{m} \pi_j g(j)$$

5 Queueing Theory

5.1 Queueing System

Components of a queueing system:

1. *Arrival:* Customers enter the queueing system and join a queue.
2. *Queue discipline:* A member of the queue is selected for service by some rule.
3. *Service:* The customer is served before leaving the queueing system.

Main uncertainty drivers are *interarrival times* (times between the arrival of customers) and *service times* (times for service of a customer).

The number of customers in the system at time t, $N(t)$, is called the *state of the system* at time t.

Kendall's notation: Simple queueing systems are conventionally labelled by:

- U: interarrival distribution
- V: service time distribution (most commonly used notation: D for deterministic, M for exponential, and G for general distribution)
- s: number of servers
- κ: capacity (maximum number of customers in service and queue)
- W: queueing discipline (e.g., FIFO, LIFO, priority based)

5.2 Arrival and Service Processes

The *arrival rate* λ is the expected number of arrivals per unit time. The expected time between two arrivals is $1/\lambda$.

The *service rate* μ is the expected number of customers that can be served by one of the servers per unit time. The expected service time per customer is $1/\mu$.

In a *Poisson arrival process* with arrival rate λ, the probability for k customers to arrive in $[0,t]$ is:

$$P(X=k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

The sequence of *interarrival times* T_n denotes the elapsed time between the $(n-1)$-th and the n-th event.

In an *exponential service process* with service rate μ, the probability density function for service time is:

$$f(t)=\mu e^{-\mu t}$$

The probability that the service time T is less than some t is

$$P(T\le t)=1-e^{-\mu t}$$

If T is exponentially distributed, then T has the lack of memory property:

$$P(T>t+s\mid T>s)=P(T>t)$$

The minimum property: If $T_1,\dots,T_n$ are independent and exponentially distributed random variables with parameters $\lambda_1,\dots,\lambda_n$ then $T=\min\{T_1,\dots,T_n\}$ is exponentially distributed with parameter $\lambda=\lambda_1+\dots+\lambda_n$.

Disaggregation of arrival processes: Consider customers arriving with exponential interarrival times at rate λ. There are n types of customers with p_i being the probability that a customer of type i arrives. Then the interarrival times of customers of type i are exponentially distributed with parameters $\lambda_i=p_i\lambda$.

Utilization:

$$\rho=\frac{\lambda}{s\mu}$$

5.3 Steady-State Metric

After sufficient time has elapsed, the state of the system becomes essentially independent of the initial state and the elapsed time.

P_n is the probability of exactly n customers in the system.

The expected number of customers in the system is:

$$L=\sum_{n=0}^{\infty}nP_n$$

The expected queue length (excludes customers being served) is:

$$L_q=\sum_{n=s}^{\infty}(n-s)P_n$$

W is the expected waiting time in the system (includes service time) for each individual customer.

W_q is the expected waiting time in the queue (excludes service time) for each individual customer.

$$W=W_q+\frac{1}{\mu}$$

Little's Formula: In steady state the expected queue length when a customer arrives is the same as the expected queue length when leaves the queue:

$$L_q=\lambda W_q,\quad L=\lambda W$$

If the system is in steady state, then the distribution characterized by the probability mass function $p_n=P(N=n)$ is called the *steady-state distribution*.

$$p_n=\lim_{t\to\infty}P(N(t)=n)$$

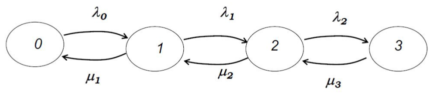
5.4 Birth-and-Death Processes

Given time t and state $N(t)=n$, the probability distribution of the remaining time until the next birth or death is exponential with parameter λ_n and μ_n respectively.

The random variables are mutually independent and the process is in steady state.

Define E_n and L_n to be the rate (average number of events per unit time) at which the system enters and leaves state n, respectively.

$$E_n=L_n$$



p_n can be interpreted as the proportion of time the process is in state n.

$$E_n=p_{n-1}\lambda_{n-1}+p_{n+1}\mu_{n+1}$$
$$L_n=p_n(\lambda_n+\mu_n)$$

Applying equality for every state:

$$\mu_1p_1=\lambda_0p_0$$
$$\mu_{n+1}p_{n+1}=\lambda_np_n,\quad n\ge 1$$

Steady-state probabilities:

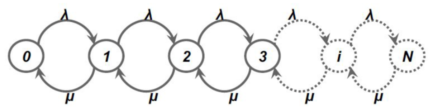
$$p_n=c_np_0,\quad c_n=\frac{\lambda_0\lambda_1\dots\lambda_{n-1}}{\mu_1\mu_2\dots\mu_n}$$

Notice that $p_0+p_1+p_2+\ldots=1$:

$$p_0=\frac{1}{1+c_1+c_2+\cdots}$$
$$p_n=\frac{c_n}{c_0+c_1+c_2+\cdots}$$

5.5 Queues

$M/M/1$: the arrival process is Poisson, the service time is of an exponential distribution, and there is only one server.



For the $M/M/1$ system, we have $\lambda_n=\lambda,\mu_n=\mu$ for all n.

$$c_n=\left(\frac{\lambda}{\mu}\right)^n=\rho^n$$
$$p_0=\frac{1}{1+\rho+\rho^2+\ldots}=1-\rho$$
$$p_n=c_np_0=(1-\rho)\rho^n$$

Average number of customers in the system, L:

$$L=\sum_{i=0}^{\infty}ip_i=\sum_{i=0}^{\infty}i\rho^i(1-\rho)=\frac{\rho}{1-\rho}$$

Average queue length, L_q:

$$L_q=\sum_{i=1}^{\infty}(i-1)p_i=\frac{\rho}{1-\rho}-(1-p_0)=\frac{\rho^2}{1-\rho}$$

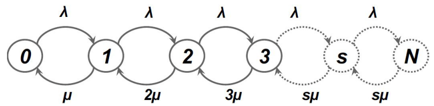
Average time customers spend in the system, W:

$$W=\frac{\rho}{(1-\rho)\lambda}=\frac{1}{\mu(1-\rho)}$$

Average waiting time, W_q:

$$W_q=W-\frac{1}{\mu}=\frac{\rho}{\mu(1-\rho)}$$

$M/M/s$: the arrival process is Poisson, the service time is of an exponential distribution, and there are s servers.



For the $M/M/s$ system, $\mu_n=n\mu$ for $n<s$ and $\mu_n=s\mu$ for $n\ge s$.

$$p_0=\frac{1}{\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!}+\frac{(\lambda/\mu)^s}{s!}\left(\frac{s\mu}{s\mu-\lambda}\right)}$$

$$p_n=\begin{cases}\frac{(\lambda/\mu)^n}{n!}p_0 & \text{if } 0\le n\le s\\ \frac{(\lambda/\mu)^n}{s!s^{n-s}}p_0 & \text{if } n\ge s\end{cases}$$

$$L_q=\left(\frac{(\lambda/\mu)^{s+1}}{(s-1)!(s-\lambda/\mu)^2}\right)p_0$$

6 Regression Analysis

6.1 Linear Regression Model

The relationship/equation that describes how the *dependent variable* y is related to an *independent variable* x and an *error term* $\varepsilon\sim\mathcal{N}(0,\sigma^2)$ is called the *regression model*.

$$y=\alpha+\beta x+\varepsilon$$

α and β are parameters of the model, called the intercept and the slope, respectively.

Curve fitting: Find parameters a and b such that the deviations between model predictions $a+bx_i$ and observations y_i are small at all data points. A popular criteria is the error sum of the squares (ESS).

Forecast errors or *residual errors* or deviations for an example line:

$$e_i=a+bx_i-y_i$$

6.2 Least Squares Method

Least squares criterion:

$$\min_{a,b}ESS(a,b)=\sum_{i=1}^{n}(a+bx_i-y_i)^2$$

Optimality conditions: Least square solution when $\frac{\partial ESS(a,b)}{\partial a}=\frac{\partial ESS(a,b)}{\partial b}=0$.

$$a+b\overline{x}=\overline{y}$$

Hence the line of best fit $y=a+bx$ passes through the *average point* $(\overline{x},\overline{y})$.

$$\overline{x}=\frac{\sum x_i}{n};\quad \overline{y}=\frac{\sum y_i}{n}$$

The optimal parameters are given by:

$$b=\frac{\sum x_i(y_i-\overline{y})}{\sum x_i(x_i-\overline{x})}$$
$$a=\overline{y}-b\overline{x}$$

6.3 Correlation

The correlation coefficient R is a measure of the strength of the linear relationship between two variables.

$$R=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

The variance and the covariance of X and Y are given by:

$$S_{XX}=\sum_{i=1}^{n}(x_i-\overline{x})^2;\quad S_{YY}=\sum_{i=1}^{n}(y_i-\overline{y})^2$$
$$S_{XY}=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$$

The bigger $|R|$ is, the stronger linear relationship between the two variables is.

6.4 Error Decomposition

Partition of the total sum of squares: Total sum of squares (TSS) = error sum of squares (ESS) + regression sum of squares (RSS).

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$$

The R^2 statistic (the coefficient of determination) indicates how well an estimated regression function fits the data. It measures the proportion of the total variation in the dependent variable y around its mean that is accounted for by the independent variable in the estimated regression equation.

$$R^2=\frac{RSS}{TSS}=1-\frac{ESS}{TSS}=\frac{S_{xy}^2}{S_{xx}S_{yy}}$$

6.5 Statistics and Regression

In statistics, *confidence interval* estimates replace point estimates so that we obtain more reliable estimates for population means and population proportions.

For a standard normal distribution (or a t-distribution), the middle area in the graph below is the confidence level and the area of the two tails is the significance level:

In statistics, we conduct *hypothesis testing* to test theories and claims based on the data.

Framework for conducting hypothesis testing (the population mean):

1. Assume m is the sample mean and μ is the population mean.
2. Make a *null hypothesis* H_0: $m=\mu$, and an *alternative hypothesis* H_1: $m\neq\mu$.
3. Select a *significance level* α and find out its critical value z.
4. Calculate test statistic: $t=(m-\mu)/SE$, where $SE=\sigma/n^{1/2}$.
5. Reject H_0 if the absolute value of t is larger than the z-value. Otherwise, do not reject H_0.

In regression analysis, the t-statistic for the slope is calculated based on the null hypothesis that the slope is equal to zero and p-value is equal to the tail probability. For a regression equation to be meaningful or significant, we want to see a large value of t-statistic and a small value of p value for the slope. Equivalently, we want to see that zero is outside of some relevant confidence interval for the slope.

The *standard error* measures the scatter in the actual data around the estimate regression line. It is an unbiased estimated value for the parameter σ, the variance of the error term.

$$S_e=\sqrt{\frac{\sum_{i=1}^{n}(y_i-(a+bx_i))^2}{n-2}}$$

It can be proved that $(b-\beta)/S_b$ is of a Student's t-distribution with $n-2$ degrees of freedom.

$$S_b=S_e/\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

7 Forecasting

7.1 Multiple Linear Regression

Multiple linear regression relates a dependent variable y to more than one independent variables $\{x_1,x_2,\ldots,x_k\}$ and the noise e which is of a normal distribution that is independent of independent variables. The multiple linear regression model is:

$$y=\alpha+\beta_1x_1+\beta_2x_2+\ldots+\beta_kx_k+e$$

α represents the intercept and β's are the coefficients of the contribution by the independent variables.

If two independent variables correlate strongly with each other, they are nearly *collinear* and we should not include both as *explanatory variables*.

Multiple regression process:

1. *Backward elimination:* Starting with all variables included, test each of these variables for a significant relationship with the dependent variable and remove the variable with the largest p-value.
2. *Forward selection:* Start with one or more variables included, test each of these variables for a significant relationship with the dependent variable Add the variable with a small p-value.
3. *Stepwise selection:* A combination of both forward selection and backward elimination.

Spurious correlation: Correlation does not imply causation. Two variables are statistically correlated, but have no causal connection in reality.

Simpson paradox: The correlation coefficient sign for the whole data set is different from the correlation coefficient signs for two or more subsets.

7.2 Time Series Forecasting

Given historical observations over time of some quantity of interest (a random variable), we study typical statistical properties:

1. *Trend:* Long term direction.
2. *Seasonality:* Repeating patterns.
3. *Cycle:* Economic cycles.
4. *Correlations:*

We predict future based on historical observations.

Moving Average: The general formula for m-period moving average is:

$$F_t=\frac{X_{t-m}+\ldots+X_{t-2}+X_{t-1}}{m}$$

We define $e_t=x_t-F_t$ as the forecast error in period t.

Measures of forecast error over n periods:

1. *Mean Absolute Deviation* (MAD):
$$MAD=\frac{\sum_{i=1}^{n}|e_i|}{n}$$
2. *Mean Squared Error* (MSE):
$$MSE=\frac{\sum_{i=1}^{n}e_i^2}{n}$$
3. *Mean Absolute Percentage Error* (MAPE):
$$MAPE=\frac{\sum_{i=1}^{n}|e_i|/x_i}{n}$$

Exponential smoothing: The smoothed average of period t is equal to the weighted average of the current observation and the smoothed average in the most recent past time period.

$$\begin{aligned}E_t&=\alpha X_t+(1-\alpha)E_{t-1}\\&=E_{t-1}+\alpha(X_t-E_{t-1})\end{aligned}$$

$\alpha\in(0,1)$ is called the smoothing parameter. The prediction equation (k-step) is $F_{t+k}=E_t$.

The weights in exponential smoothing:

$$E_t=(1-\alpha)^{k+1}E_{t-k-1}+$$
$$\alpha\left[X_t+(1-\alpha)X_{t-1}+\ldots+(1-\alpha)^kX_{t-k}\right]$$

The forecast of exponential smoothing takes all past observations into account by giving more weights on more recent observations (the weight decreases exponentially back in time).

Winters' multiplicative smoothing:

1. *Base:*
$$E_t=\alpha\frac{X_t}{S_{t-c}}+(1-\alpha)(E_{t-1}+T_{t-1})$$
2. *Trend:*
$$T_t=\beta(E_t-E_{t-1})+(1-\beta)T_{t-1}$$
3. *Seasonality:*
$$S_t=\gamma\frac{X_t}{E_t}+(1-\gamma)S_{t-c}$$

Prediction equation (k-step):

$$F_{t+k}=(E_t+kT_t)S_{t+k-c}$$

$\alpha,\beta,\gamma\in(0,1)$ are given smoothing parameters. c is the number of seasons.

Winters' additive smoothing:

$$\begin{aligned}E_t&=\alpha(X_t-S_{t-c})+(1-\alpha)(E_{t-1}+T_{t-1})\\T_t&=\beta(E_t-E_{t-1})+(1-\beta)T_{t-1}\\S_t&=\gamma(X_t-E_t)+(1-\gamma)S_{t-c}\end{aligned}$$

Here seasonality is used as an addition rather than a multiplier.

$$F_{t+k}=E_t+kT_t+S_{t+k-c}$$

Exponential smoothing with trend:

$$\begin{aligned}E_t&=\alpha X_t+(1-\alpha)(E_{t-1}+T_{t-1})\\T_t&=\beta(E_t-E_{t-1})+(1-\beta)T_{t-1}\end{aligned}$$

$$F_{t+k}=(E_t+kT_t)$$

Exponential smoothing with multiplicative seasonality:

$$\begin{aligned}E_t&=\alpha\frac{X_t}{S_{t-c}}+(1-\alpha)E_{t-1}\\S_t&=\gamma\frac{X_t}{E_t}+(1-\gamma)S_{t-c}\end{aligned}$$

$$F_{t+k}=E_tS_{t+k-c}$$

Exponential smoothing with additive seasonality:

$$\begin{aligned}E_t&=\alpha(X_t-S_{t-c})+(1-\alpha)E_{t-1}\\S_t&=\gamma(X_t-E_t)+(1-\gamma)S_{t-c}\end{aligned}$$

$$F_{t+k}=E_t+S_{t+k-c}$$

7.3 Confidence Intervals

The square root of MSE is the standard deviation of the forecast errors.

Assuming that the forecast errors are normally distributed and that past trends continue into the future, there is a 95% chance that a future value, when it occurs, will be within approximately two standard deviations of the point estimate made in the time series forecasting method.

8 Portfolio Management

8.1 Markowitz Portfolio Analysis

The *return rates* are uncertain numbers with mean (% per week) and variance σ^2.

We can invest a proportion α between 0 and 1 of our money into D, and the rest into N. Then we form a portfolio v such that the uncertain portfolio value is:

$$v=\alpha D+(1-\alpha)N$$

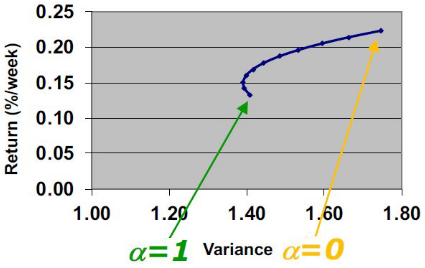
The expected return of portfolio v is:

$$r_v=\alpha r_D+(1-\alpha)r_N$$

The variance of portfolio v is:

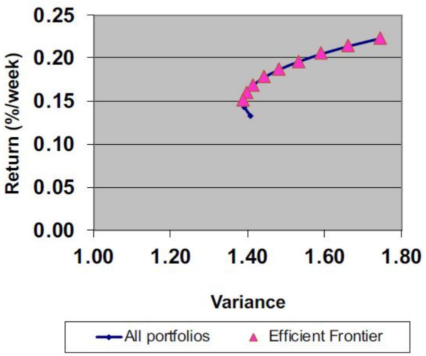
$$\begin{aligned}\sigma_v^2=&\ \alpha^2\sigma_D^2+(1-\alpha)^2\sigma_N^2\\&+2\alpha(1-\alpha)[\mathbb{E}(DN)-r_Dr_N]\end{aligned}$$

Mean-variance diagram:



8.2 Efficient Frontier

Portfolio is *efficient* if you cannot improve both return and risk at the same time.



Given any number $x_1, \ldots, x_n$ of uncertain numbers and any proportions $\alpha_1, \ldots, \alpha_n$ (non-negative and sum to 1) and corresponding portfolio with value:

$$v = \alpha_1 x_1 + \ldots + \alpha_n x_n$$

Given mean r_i of each x_i, mean portfolio value is:

$$r_v = \mathbb{E}(v) = \alpha_1 r_1 + \ldots + \alpha_n r_n$$

Risk of portfolio value, measured in variance, is:

$$\sigma_v^2 = \operatorname{var}(v) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \operatorname{covar}\left(x_i, x_j\right)$$

The efficient frontier can be found by minimizing $\operatorname{var}(v)$ subject to $r_v \geq r$ and $0 \leq \alpha_1, \ldots, \alpha_n \leq 1$.

8.3 Hedging

Hedging relies on negative correlation (statistical dependence) between x and y. Hedging reduces risk with few investments.

Diversification relies many/several different investments that have some statistical independence. Diversification is robust to positive dependence.

8.4 Covariance and Variance of Portfolios

Variance is not an ideal measure of risk. It describes the risk of not being average. *Semi-variance* describes the risk of loss. Lower semi-variance:

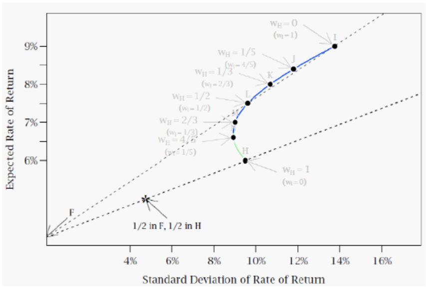
$$\operatorname{semi-var}(v) = \mathbb{E}\left(\min\left(v - r_v, 0\right)^2\right)$$

Semi-variance equals $1/2$ of variance if distribution is symmetric.

8.5 Market Portfolio

If a riskless asset is added to the portfolio, then investors can borrow and lend money. Let r_f be the riskless rate of return.

Then any point on the line that connects a portfolio and the riskless asset forms a new portfolio. Such a line is called *capital allocation line*.



Let α = proportion allocated to risky portfolio, P, and $1 - \alpha$ = proportion allocated to risk-free asset, F.

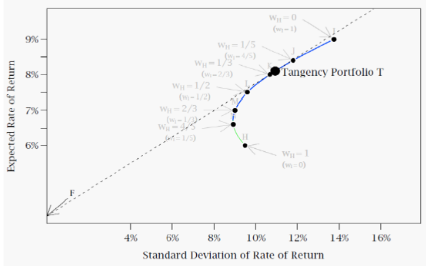
The expected return of the complete portfolio is:

$$\begin{aligned} \mathbb{E}(r_v) &= \alpha\left[\mathbb{E}(r_P)\right] + (1 - \alpha) r_F \\ &= r_F + \alpha\left[\mathbb{E}(r_P) - r_F\right] \end{aligned}$$

Since $\sigma_v = \alpha \sigma_P$, then $\alpha = \frac{\sigma_v}{\sigma_P}$. The equation of the capital allocation line is:

$$\mathbb{E}(r_v) = r_F + \frac{\sigma_v}{\sigma_P}\left[\mathbb{E}(r_P) - r_F\right]$$

The *capital market line* (CML) is the one with the highest slope among all capital allocation lines. The corresponding portfolio on the efficient frontier is called the *market portfolio*.



Any portfolio on the original efficient frontier is dominated by some portfolio on the capital market line.

Let $\left(r_f, 0\right)$ be the risk-free asset and (r_M, σ_M) be the market portfolio. A portfolio (r_v, σ_v) on the capital market line satisfies the following equation:

$$r_v - r_f = \frac{r_M - r_f}{\sigma_M} \sigma_v$$

8.6 Risk of an Asset

Total risk = specific risk + systematic risk:

1. *Specific risk* is also known as unsystematic, idiosyncratic risk. It is firm-specific and can be diversified away when you hold a number of assets.
2. *Systematic risk* is market-wide conditions and not possible to diversify away.

Beta for an asset is used to measure the risk that a security contributes to a diversified portfolio.

$$\beta_i = \frac{\operatorname{covar}(i, M)}{\sigma_M^2} = \frac{\operatorname{covar}(i, M)}{\sigma_M \sigma_i} \frac{\sigma_i}{\sigma_M}$$

Beta for an asset can be obtained by conducting simple regression analysis for this asset (dependant variable) and the market portfolio.

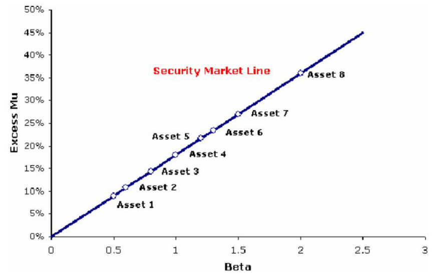
8.7 Capital Asset Pricing Model (CAPM)

The capital market line relates the expected rate of return of an *efficient portfolio* to its standard deviation, but it does not show the expected rate of return of an *individual asset* relates to its individual risk.

Capital Asset Pricing Model (CAPM): If the market portfolio M is efficient, then the expected rate of return r_i of any asset i satisfies:

$$r_i - r_f = \beta_i \left(r_M - r_f\right)$$

The *Security Market Line* (SML) is the line of Beta vs. excess return $r_i - r_f$ of assets with a slope of $r_M - r_f$. It tells us the reward for bearing a certain level of risk.



(The End)