3E3 Modelling Risk

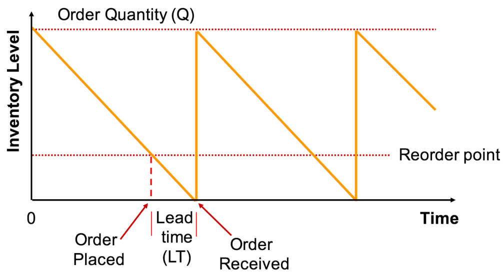
'Need to Know...'

Mrinank Sharma

April 4, 2018

This is an attempt to condense everything which **must be memorised** for this course and hence does not include data-book content.

1. Under deterministic demand, the inventory cycle follows a sawtooth. *The Inventory Cycle*



The cost per period is given by *Cost per Period*

$$T(Q) = \frac{Q}{2}C_h + \frac{D}{Q}C_O \Rightarrow \text{EOQ} = \sqrt{2D\frac{C_O}{C_H}}$$

where Q is the batch size, D is the demand per period, C_O is the order cost and C_H is the holding cost. The Economic Order Quantity is determined by using the first order condition. Strictly speaking, the holding cost for safety stock and the variable cost per item should be included.

Problems with this model is that the assumptions made are very rigid (e.g. many quantities are assumed to be constant). It also relies on getting good number of the holding and ordering costs. It is robust, tends to inflate order sizes. Empirically, $< 12\%$ from the optimum.

2. The Newsvendor Model applies under stochastic demand. The demand CDF is required as are the overage/underage costs. *Newsvendor Model*

$$G(Q, D) = c_o \max(0, Q - D) + c_u \max(0, D - Q)$$
$$\frac{d}{dQ}\left(\mathbb{E}_D[G(Q, D)]\right) = 0 \Rightarrow F_D(Q^*) = \frac{c_u}{c_o + c_u}$$

and thus it is easily see that if $c_u = c_o$, $Q^* = \mathbb{E}[D]$

3. This is an extension of the EOQ model but now both Q and R are decision variables and the cycle time is not constant. Demand is assumed to be random and stationary with a fixed lead time, L. Denote the holding cost as C_H, the setup cost per order C_O and the penalty per unit of unsatisfied demand p. The expected demand per unit time is λ. *(Q,R) Policy*

$$C(Q) = C_h(\underbrace{\frac{Q}{2} + R - \lambda L}_{\text{Average Stock}}) + \frac{C_O}{Q/\lambda} + p\frac{\mathbb{E}[\max(0, D - R)]}{Q/\lambda}$$

The first order conditions can be applied to Q and R leading to the following equations which must be solved iteratively

$$Q = \sqrt{\frac{2\lambda(C_O + pn(R))}{C_H}} \qquad F_D(R) = 1 - \frac{QC_H}{p\lambda}$$

The final solution is a compromise between the shortage costs (large R), holding costs (small Q and R) and the fixed costs.

4. There are the number of decisions rules which can be used: *Decision Rules*

   - **Minimax**: Maximise the absolute maximum possible payoff. Generally highly risky.
   - **Maximin**: Maximise the minimum possible payoff. Quite conservative.

5. Often it the probabilities of different states of nature can be determined. *Decision Analysis: Definitions*

   (a) The expected monetary value is the expected value of decision i

   $$\mathsf{EMV}_i = \sum_j r_{ij} p_j$$

   where r_i is the payoff for decision i under outcome j.

   (b) The expected opportunity loss is the expected regret of decision i

   $$g_{ij} = \underbrace{(\max_k r_{kj}) - r_{ij}}_{\text{Regret}} \qquad \mathsf{EOL}_i = \sum_j g_{ij} p_j = \sum_j (\max_k r_{kj}) p_j - \mathsf{EMV}_i$$

   i.e. the regret is the difference between the payoff of the decision chosen and the maximum payoff given an outcome. Thus, **the decision with the largest EMV has the minimum regret**.

   (c) If the future could be predicted with 100% accuracy, the strategy could be optimised.

   $$\mathsf{PI} = \sum_j (\max_i r_{ij}) p_j$$

   The expected value of perfect information is the difference between PI and the maximum EMV (which is the same as the minimum EOL).

   $$\mathsf{EVPI} = \mathsf{PI} - \max_i(\mathsf{EMV}_i) = \min(\mathsf{EOL}_i)$$

   (d) If probabilities estimates can be refined, they add value. The **expected value of sample information** (EVSI) is the difference between the EMV with the same information and the EMV without.

These measures don't consider the spread of pay-offs, an issue remedied by utility theory. Replacing the pay-offs with the utility for a payoff is a solution to this (and the utility curve defines the risk preferences e.g. $\mathbb{E}[U(X)] > U(\mathbb{E}[X])$ describes risk seeking preferences).

6. Dynamic programming is a technique used to solve complex problems by breaking them down into smaller problems and then solving using *backwards induction*. The problem is defined as follows; note that a * denotes optimality. *Dynamic Programming*

   - N is the number of stages.
   - $n \in \{1, \ldots, N\}$ is the stage index.
   - s_n is the state variable for stage n.
   - s_n is the decision variable for stage n.
   - $s_{n+1} = g_n(s_n, x_n)$ i.e. $g(\cdot)$ represents the system transition between stages given a choice of the decision variable.
   - $c_n(s_n, x_n)$ represents the contribution to the system value in a given stage, n.
   - $f_n(s_n, x_n)$ represents the *total* contribution of all following stages if the state is s_n and the decision taken is x_n and optimal decisions are made subsequently.

Thus the Bellman optimality equation is formed *Bellman Equation*

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

This technique finds use in stochastic situations where the transition function, $g(\cdot)$, is replaced by a probability mass function and the expectation is minimised (in the Bellman equation).

7. A few definitions for markov chains have been included below: *Markov Chains*

    - State j is accessible from i if there is a path from $i \to j$. If $i \leftrightarrow j$, i & j are said to communicate.
    - A class, A, is absorbing if for each $i \in A$, $i \to j \Rightarrow j \in A$. Thus it is impossible to escape from an absorbing class. A state is absorbing if it is a absorbing class itself.
    - Class B is accessible from class A if $\forall i \in A$ & $\forall j \in B, i \to j$.
    - A chain with a single class is irreducible.
    - A state is recurrent can be visited an infinite number of times. A transient state is only visited a finite number of times.
    - A state is aperiodic if $p_{ii}^(n) > 0$ for all sufficiently large n. Note that this is a class property.
    - A Markov chain is period with period m if the process can only recur to state i after km steps.
    - If a Markov chain is irreducible and aperiodic then there is a limiting distribution.
    - The invariant distribution is given by $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$; this is useful to calculate e.g. average payoff per unit time.

    Note that, by the law of total probability,

    $$p_{ij}^{(n)} = \sum_k p_{jj}^{(n-k)} f_{ij}(k)$$

    where $p_{ij}^{(n)}$ is the probability of ending in state j after n steps after starting in state i and $f_{ij}(k)$ is the probability that the first passage from i to j occurs after k transitions. This gives a recursive set of relations. This allows the expected first passage time to be calculated, $\sum_k k f_{ij}(k)$. Alternatively, a set of linear equations can be set up.

8. With continuous time Markov chains, the rates determine the density function of the waiting times: *Continuous Time Markov Chains*

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

    The probability of arriving in an interval is $\lambda \Delta t$. The exponential distribution shows the *lack of memory property* as required. In order to find the steady state, consider the *balance equations* and recall that the probabilities must sum to one.

9. Simple queueing systems are labelled as $U/V/s/\kappa/W$ where *Queueing Notation*

    - U, V denote the inter-arrival and service time distributions. Can be D deterministic, M memoryless or G general.
    - s denotes the number of serves.
    - κ is the system capacity.
    - W is the queueing protocol (e.g. FIFO).

    Define the utilization factor, $\rho = \frac{\lambda}{s\mu}$, which is the expected fraction of time the service is expected to be busy. We define the follow additional performance measures where p_i is the probability of having i customers:

    (a) Expected number of customers, $L = \sum_i^\infty ip(i)$.
    (b) Expected queue length of a system with s servers, $L_q = \sum_i^\infty ip(i+s)$.
    (c) Average queue waiting time, W_q.
    (d) Average waiting time, $W = W_q + \frac{1}{\mu}$.

    Littles formula says that $L_q = \lambda W_q$ and $L = \lambda W$ **at steady state** since the number of customers leaving the queue must also enter the queue in steady state.

10. If $X_1, \ldots, X_n$ are i.i.d. according to the exponential distribution, the minimum is also exponentially distributed with parameter $\lambda = \sum_i \lambda_i$. When there are multiple customer types, the average inter-arrival time is taken to be the mean.

11. We assume that $y = \alpha + \beta x + \epsilon$ and see to find parameters such that the deviations between model predictions and observations is small $\hat{y}_i = a + bx_i$. This is done by minimising the least squares error using the familiar first order conditions. *Regression*

$$\text{ESS} = \text{S} = \sum_i (y_i - \hat{y}_i)^2 \Rightarrow b = \frac{S_{XY}}{S_{XX}}, a = \bar{y} - b\bar{x}$$

$$\text{where } S_{XX} = \sum_i (x_i - \bar{x})^2 \qquad S_{XY} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

There are also other forms of deviation including the total sum of squares, TSS $= \sum_i (y_i - \bar{y})^2$ which measures the variation in y around the mean and the regression sum of squares, RSS $= \sum_i (\hat{y}_i - \bar{y})^2$. Note that TSS $=$ ESS $+$ RSS.

It is important to understand the summary statistics:

- $R = \frac{S_{XY}}{\sqrt{S_{XX} S_{XY}}}$ is known as the **correlation coefficient** and is a measure of the strength of the relationship. Note that $b = R\sqrt{\frac{S_{YY}}{S_{XX}}}$.
- $R^2 = \frac{RSS}{TSS}$ (where the square is only valid for regression); the closer to 1, the better the estimated regression fits the data. There also exists an *adjusted* R^2 statistic which accounts for the number of data points.
- The standard error, $S_e = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-1}}$, measures the scatter in the data around the regression line.
- A rule of thumb for approximating a 95% prediction is $[\hat{y} - 2S_e, \hat{y} + 2S_e]$.
- Large values of the t-statistic and small values of the p-value indicate that a variable used is significant. The p-value represents the trail probability of the variable and is used with normal hypothesis testing.

a and b are unbiased estimators which are in fact normally distributed which allows confidence values to be obtained. There is a problem of multicollinearity in multiple regression models if two variables used are highly correlated. Coefficient estimates may change dramatically when a variable is removed. This can give large p-values even if the variable used is significant. Note that the model can be easily extended to non-linear regression.

The forecasting error is defined as the difference between actual and forecasted value. There are several measures of forecast accuracy including: *Forecasting*

- Mean absolute deviation, MAD $= \frac{1}{N} \sum_i |e_i|$.
- Mean squared error, MSE $= \frac{1}{N} \sum_i e_i^2$.
- Mean Absolute Percentage Error MAPE $= \sum_i |\frac{e_t}{X_t} \frac{100}{N}|$

There are many different forecasting methods:

(a) The Moving Average

$$F_{t+1} = \frac{X_t + \cdots X_{t-m+1}}{m}$$

(b) Holt-Winter's Multiplicative Smoothing (**Don't Learn - on the datasheet!**)

$$E_t = \alpha \frac{X_t}{S_{t-c}} + (1 - \alpha)(E_{t-1} + T_{t-1})$$
$$T_t = \beta(E_t - E_{t-1}) + (1 - \beta)T_{t-1}$$
$$S_t = \gamma \frac{X_t}{E_t} + (1 - \gamma)S_{t-c}$$
$$F_{t+k} = (E_t + kT_t)S_{t+k-c}$$

c is the number of time periods per season. The parameters can be set to zero also to give simpler forms of smoothing e.g. $T = 0$ gives exponential smoothing with seasonality. To initialise this method, a complete season of data is required to determine the initial estimates for S_t (but it is better to have two seasons for the trend).

(c) Holt-Winter's Additive Smoothing

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

In each case, the best values of the parameters can be found via optimisation. Suitable initialisations for the values include $E_0 = X_0$. Set T_0 to the average trend across a season and $S_{i-c} = \frac{X_{i-c}}{\sum_j^c X_{j-c}/c}$.

12. An investor choosing a portfolio, $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ which are real numbers (thus assuming **divisibility, liquidity and allowing for shorted assets**). Define *Portfolio Management*

- Assume $A_i(t)$ is the wealth at time t of the asset i. The wealth is

$$V(t) = \sum_i \alpha_i A_i(t)$$

- The rate of return is

$$K_A = \frac{A(t) - A(0)}{A(0)}$$

- The portfolio can be expressed in weights. Thus the portfolio return is

$$K_v = \sum_i \omega_i K_i \text{ where } \omega_i = \frac{\alpha_i A_i(0)}{V(0)}$$

Note that the prices are stochastic; we seek to maximise the expectation of the rate of return. Risk is quantified using the standard deviation. It is possible to plot portfolios on a graph of mean and variance; investors will then choose the minimum variance portfolio for a given mean (or the largest mean for a given variance). A portfolio is efficient if there is no other portfolio which has a strictly higher return with a strictly lower variance.

Many models treat portfolio selection as a constrained optimisation problem e.g. minimise the variance given a mean. These problems can be solved using many software packages. Note as the number of securities increases, the optimisation problem becomes harder.

A risk-less asset has zero standard deviation.Any point on the line that connects a portfolio and the risk-less asset forms a new portfolio. In this case, any efficient portfolio can be expressed as a combination of P and the risk-less asset. The **one-fund theorem** states that there is a single fund of risky assets P that can be decomposed as a combination of the fund P and the risk-free asset.

In market equilibrium, it is assumed that everybody solves the same problem and thus every investor holds the **market portfolio**. The weight is equal to the proportion of that asset's total capital value over the total market cap. Thus the optimal portfolio is the market portfolio and we don't need to solve the problem. Equilibrium in the market will force the market into the direction that solves the mean-variance problem since the prices will change, thus changing the optimal portfolio. This is similar to supply and demand matching.

The capital market line is the line with the highest slope amongst all lines connecting the risk-less asset to the curve (and the tangency point is the market portfolio). This dominates any portfolio on the original frontier and all portfolios on this line have the same incremental rate of return per unit risk. Thus for any efficient portfolio:

$$r = r_f + \frac{r_M - r_F}{\sigma_M}\sigma$$

The β for an asset measures the risk that the security contributes to a portfolio. If the market portfolio M is efficient then

$$r_i - r_f = \beta_i(R_m - r_f) \qquad \beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$$

$r_i - r_f$ is the expected excess rate of return of the asset which is linked to the excess rate of return of the market portfolio with proportionality factor β.