INSE6290	Jia Yuan Yu
3: Uncertainty, Newsboy, Forecasting	
Concordia	October 14, 2016

1 What is uncertainty?

Everything that is not certain. Analogy:

- What you know;
- What you don't know, and what you don't know that you don't know.

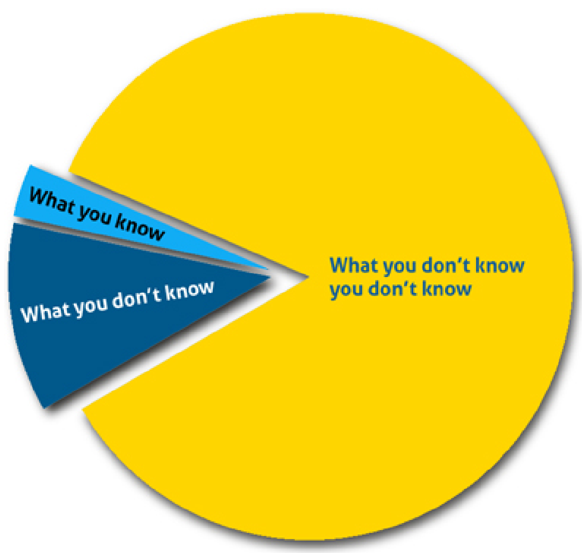

Figure 1: From http://www.tecplot.com/wp-content/uploads/2012/09/dontknow.jpg

The most widely used mathematical model for uncertainty uses probability.

1.1 Review of probability

- Probability space,
  - Set of outcomes Ω,
  - Set of events $\mathcal{F}$ containing subsets of Ω,
  - A function (probability measure) $P : \mathcal{F} \to \mathbb{R}$.
- Random variables, distribution functions,

- Real-value random variable is a (measurable) function $X : \Omega \to \mathbb{R}$.
- Probability distribution function or cumulative distribution function: $F(x) = P(X \leq x)$ for all x.
- Probability density function, if it exists: $F(x) = \int_{-\infty}^{x} f(z)dz$.

* Examples (Bernoulli, Normal, Uniform, Exponential etc.),
  - Bernoulli with parameter p: $P(X = 1) = p$, $P(X = 0) = 1 - p$.
  - Exponential with rate λ: $F(x) = (1 - e^{-x})1_{[x \geq 0]}$ for all $x \in (-\infty, \infty)$.
  - Normal $N(0, 1)$: $f(x) = (2\pi)^{-1/2} e^{-x^2/2}$ for all $x \in (-\infty, \infty)$.

* Expectation,
  - Bernoulli: $\mathbb{E}X = p * 1 + (1 - p) * 0$

2 Newsboy solution

As seen last week, the (random) cost for the Newsboy problem is

$$g(s, D) = h \max(s - D, 0) + p \max(D - s, 0).$$

The expected cost $\mathbb{E}g(s, D)$ is

$$\begin{aligned} \bar{g}(s) \triangleq \mathbb{E}g(s, D) &= \int_0^\infty [h \max(s - z, 0) + p \max(z - s, 0)] f_D(z)dz \\ &= \int_0^s h(s - z) f_D(z)dz + \int_s^\infty p(z - s) f_D(z)dz. \end{aligned}$$

We are looking for that optimal value of s, so let's look at the first and second derivatives of $\mathbb{E}g(s, D)$.

$$\begin{aligned} \frac{d}{ds}\bar{g}(s) &= h\frac{d}{ds}s\int_0^s f_D(z)dz - h\frac{d}{ds}\int_0^s zf_D(z)dz + p\frac{d}{ds}\int_s^\infty zf_D(z)dz - p\frac{d}{ds}s\int_s^\infty f_D(z)dz \\ &= h\frac{d}{ds}sF_D(s) - p\frac{d}{ds}s(1 - F_D(s)) - h\frac{d}{ds}\int_0^s zf_D(z)dz + p\frac{d}{ds}\left(M - \int_0^s zf_D(z)dz\right), \end{aligned}$$

where M is the constant $\int_0^\infty zf_D(z)dz$, and we used the definition of F_D. Observe that

$$\frac{d}{ds}sF_D(z) = sF'_D(s) + F_D(s) = sf_D(s) + F_D(s)$$

and that[1]

$$\frac{d}{ds}\int_0^s zf_D(z)dz = sf_D(s) - 0f_D(0) + \int_0^s 0dz = sf_D(s).$$

[1]Use Leibniz's rule to move the derivative inside the integral: $\frac{d}{ds}(\int_{a(s)}^{b(s)} g(z)dz) = g(b(s))b'(s) - g(a(s))a'(s) + \int_{a(s)}^{b(s)}[\frac{d}{ds}g(z)]dz$.

Hence, we have

$$\frac{d}{ds}\bar{g}(s) = (h+p)(sf_D(s) + F_D(s)) - p - (h+p)sf_D(s) + p \cdot 0$$
$$= (h+p)F_D(s) - p.$$

The second derivative is

$$\frac{d^2}{ds^2}\bar{g}(s) = (h+p)f_D(s),$$

which is non-negative for all $s \geq 0$. Hence, $\bar{g}$ is convex, and setting $\frac{d}{ds}\bar{g}(s) = 0$, we find that the value of s that minimizes expected cost is

$$s^* = F_D^{-1}\left(\frac{p}{h+p}\right),$$

where h is the cost per unit of excess inventory, and p is the cost per unit of inventory shortfall.

In summary, we have presented a solution to the newsboy problem that takes inputs h, p, and the distribution F_D, and outputs an optimal order size s^*. We assume that F_D is given; however, we can also estimate F_D from past observations of the demand.

Homework: How can we use the Newboy solution in beer game?

3 Forecasting

Forecasting is about time (past and future) and quantities (demand, inventory levels, etc.). In the context of supply chains, we are most interested in discrete time: $1, 2, \ldots$ The quantities of interest in the supply chains are of the form $X_1, X_2, \ldots$, taking values in $\mathbb{R}^d$. At time n, the time steps $1, \ldots, n$ are in the past and the values $X_1, \ldots, X_n$ are observations already made, whereas the time steps $n+1, n+2, \ldots$ are in the future and the values $X_{n+1}, X_{n+2}, \ldots$ are not yet seen. However, at time n, the decision-maker can make forecasts $\hat{X}_{n+1}, \hat{X}_{n+2}, \ldots$ for these values. Later in the future, at time $n+2$ say, we can compare $\hat{X}_{n+1}, \hat{X}_{n+2}$ with the true values X_{n+1}, X_{n+2}.

3.1 Pitfalls

Weather forecasts are simple compared to supply chain forecasts: a forecast of 10mm of rain has little effect on the actual realization of rain.

- Ignoring important factors in our model
  - Demand and decisions made are interconnected,
  - Marketing campaign affects demand,
  - Other participants, competitors,
- Effect of forecast on the future realization,

– Actions taken in response to the forecasts,
– Positive reinforcement: self-fulfilling effect (meeting self-imposed targets),

Responsible forecast must incorporate all the above.

3.2 Bass model

In order to forecasts to make sense, we must first make reasonable assumptions to obtain a model of the system whose quantities $X_1, X_2, \dots$ we want to forecast.

In the newsboy problem, we had the simplest model for demand: the same probability distribution F_D everyday. This is suitable for products that do not change over time: food, newspapers, etc. In this section, we consider a model for demand for products that are new. Consider a population of m customers $1, \dots, m$. For every customer i, we model the time of adoption (purchase) of the product by the customer by a random variable T taking values on $[0, \infty]$. Let $T_1, \dots, T_m$ denote the times at which customers $1, \dots, m$ adopt a given new product. We assume that these random variables are independent and identically distributed according to F_T.

The probability distribution F_T has the following parameters: the coefficient of innovation p and the coefficient of imitation q. Let f_T denote the corresponding probability density. The density function at time t is model as

$$\frac{dF_T(t)}{dt} = f_T(t) = [p + qF_T(t)](1 - F_T(t)),$$

where $p + qF_T(t)$ is the density function conditioned on the event that the customer has not yet adopted at time t, and $1 - F_T(t)$ is the probability that it was not adopted already. Notice that the imitation term $qF_T(t)$ is proportional to the fraction $F_T(t)$ of customer who already adopted the product, whereas the innovation term is not. Solving this differential equation for F with the initial condition $F(0) = 0$, we obtain[2]

$$F_T(t) = \frac{1 - e^{-(p+q)t}}{1 + (q/p)e^{-(p+q)t}}, \quad \text{for all } t \geq 0.$$

This distribution is shown in Figure 3.2. It has a density as shown in Figure 3.2.

Given the distribution F_T, we can also model other interesting quantities. The number $N_{a,b}$ of customers adopting the product between time instants a and b is

$$N_{a,b} = \sum_{i=1}^{m} 1_{[a<T_i<b]},$$

where $1_{[a<T_i<b]}$ is a Bernoulli random variable with parameters $\mathbb{P}(a < T_i < b) = F_T(b) - F_T(a)$. Finally, we conclude that the random variable $N_{a,b}$ has a binomial distribution (cf. Figure 3.2) with the following density:

$$\mathbb{P}(N_{a,b} = k) = \binom{m}{k} \Big(F_T(b) - F_T(a)\Big)^k \Big(1 - F_T(b) + F_T(a)\Big)^{m-k}.$$

[2] Homework: check this.

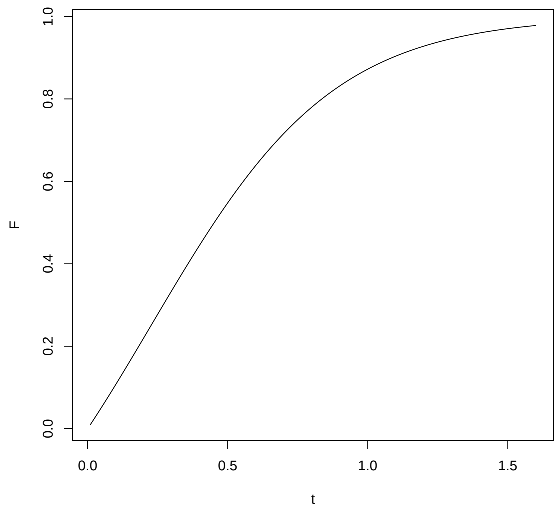

Figure 2: The distribution F_T.

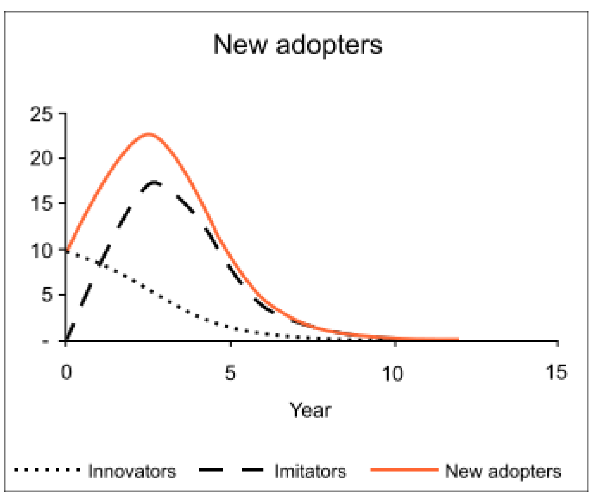

Figure 3: From https://upload.wikimedia.org/wikipedia/commons/thumb/d/d0/Bass_new_adopters.svg/339px-Bass_new_adopters.svg.png

Remark 1. From the above density, we can obtain the cumulative distribution function F_N for the demand of the new product over the time interval $[a, b]$. This F_N in turn can be used as input to a newsboy problem to find the optimal number of new products to produce or order.

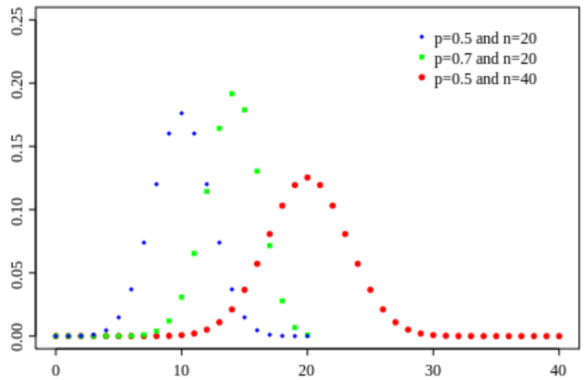

Figure 4: From https://en.wikipedia.org/wiki/Binomial_distribution

3.3 Linear regression forecast

Figure 5: From http://www.cbsolution.net/

In regression, we have a sequence of n observations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$, where each $X_j \in \mathbb{R}^d$ is a sample (e.g., number of students registered for courses 1 to d) and $Y_j \in \mathbb{R}$ is an associated quantity (e.g., how many copies of the FOSCT textbook to order). The problem is as follows: given $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ and X_{n+1} as input, forecast Y_{n+1}.

Suppose that for all $j = 1, \dots, n$:

$$Y_j = g(X_j) + \nu_j,$$

where $\{\nu_i\}$ are noise random variables that are independent and have zero mean. The function g is unknown. The regression problem is to find a function $\hat{g}_n$ using $X_1, Y_1, \dots, X_n, Y_n$ in order to estimate g.

The linear regression problem is a special case, where we assume additionally that there exists an (unknown) $\theta \in \mathbb{R}^d$ such that $g(z) = \theta^T z$[3]. In this case, the assumption on Y_j becomes:

$$Y_j = \theta^T X_j + \nu_j, \quad j = 1, \dots, n, \tag{1}$$

so that the problem becomes estimating θ by $\hat{\theta}_n$. Once we have $\hat{\theta}_n$, we can forecast Y_{n+1} with $\hat{Y}_{n+1} = \hat{\theta}_n X_{n+1}$.

3.3.1 Least-squares approach

The *least-squares* estimate of the parameter θ is

$$\hat{\theta}_n = \arg\min_{\omega \in \mathbb{R}^d} \sum_{j=1}^{n} (Y_j - X_j^T \omega)^2,$$

where

$$h(\omega) = \sum_{j=1}^{n} (Y_j - X_j^T \omega)^2$$

is convex and non-negative.

For $\hat{\theta}_n$ to be a minimum, we must have the first derivative equal to zero:

$$\frac{\mathrm{d}h}{\mathrm{d}\omega^i}(\hat{\theta}_n) = 0, \quad \text{for all } i = 1, \dots, d.$$

This is the derivative of a scalar function h by a vector[4]; hence, by the properties of the derivative[5], we obtain:

$$2 \sum_{j=1}^{n} X_j (Y_j - X_j^T \omega) = 0. \tag{2}$$

Letting

$$B \triangleq \sum_{j=1}^{n} X_j X_j^T,$$

[3] This denotes an inner product.

[4] See https://en.wikipedia.org/wiki/Matrix_calculus#Scalar-by-vector

[5] See https://en.wikipedia.org/wiki/Matrix_calculus#Scalar-by-vector_identities

the above (2) becomes

$$\sum_{j=1}^{n} X_j Y_j = B\hat{\theta}_n.$$

Hence, if B is an invertible matrix, then we have

$$\hat{\theta}_n = B^{-1} \sum_{j=1}^{n} X_j Y_j. \tag{3}$$

3.4 Moving average forecast

There is a sequence of demand values $y_1, y_2, \ldots$ with the positive integers as time indices. At time t, we have observed $y_1, \ldots, y_{t-1}$ and wish to forecast the unseen value y_t. For $N < t$, the N-th order moving average forecast is

$$\hat{y}_t = \frac{1}{N} \sum_{i=t-N}^{t-1} y_i.$$

Homework: for what assumptions and models does this work?

3.5 Exponential smoothing forecast

For $\alpha \in [0, 1]$, the exponential smooting forecast is

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1} = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i y_{t-i-1} * 1_{[t-i-1>0]}.$$

Homework: for what assumptions and models does this work?

4 Reading material

Books:

- FOSCT Section 4.4.2 and Wikipedia for Newsboy.
- For Bass model: "Why the Bass model fits without decision variables[6]" by Bass, Krishnan, and Jain, Marketing Science, 13(3), 1994.

[6]PDF