INDR 450/550

Spring 2022

Lecture 21: Prescriptive analytics 3

May 9, 2022

Fikri Karaesmen

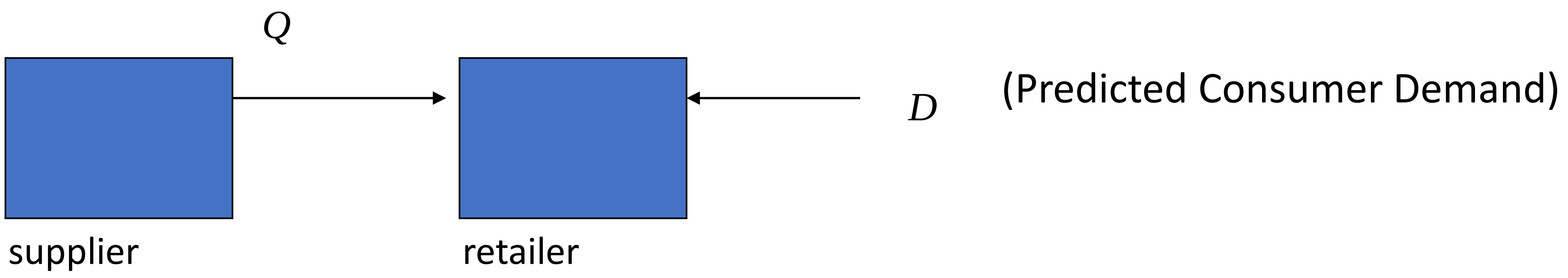
Announcements

- Class Exercise at the end of lecture today. If you are participating online, please upload your document under Course Contents/Class Exercises
- Lab 9 material (on non-linear transformations,splines are available)
- Exam scheduled for May 7 is postponed to May 13.
  - Review exercises are available
  - Make sure that you also review the class exercises and the homeworks

The Assumptions

- Short selling season
- Decision made in advance of the season
- No replenishments or capacity additions during the season (purchasing in advance is required)
- Demand forecasts have considerable uncertainty
- Items lose value significantly after the season

The Newsvendor Problem



D : demand (random variable)
Q : quantity ordered from supplier

w : wholesale price (of supplier)
r : retail price ($r>w$)
s : salvage value ($s<w$)
m: unit manufacturing cost of supplier ($m<w$)

The Newsvendor Problem

The profit as a function of Q :

$$\Pi_R(Q) = r\min(Q, D) + s(Q-D)^+ - wQ$$

Because D is a r.v., we choose to maximize:

$$E[\Pi_R(Q)] = E[r\min(Q, D) + s(Q-D)^+ - wQ]$$

$$(Q-D)^+ = \begin{cases} Q-D & \text{if } Q > D \\ 0 & \text{otherwise} \end{cases}$$

The Newsvendor Problem: the result

Solving for the optimal Q :

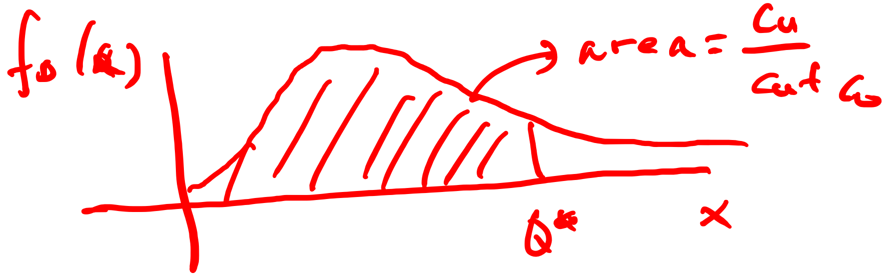
$Q^{**} = E[D]$?

$$F_D(Q^*) = \frac{c_u}{c_u + c_o}$$

$\Rightarrow Q^* = F_D^{-1}\left(\frac{c_u}{c_u + c_o}\right)$

$c_u \nearrow \quad Q^* \nearrow \qquad c_o \searrow \quad Q^*$

Q^* is such that the probability of satisfying all the demand $P(D \leq Q^*)$ is equal to the critical fraction : $c_u/(c_u+c_o)$



if $c_u = c_o \Rightarrow Q^* = F_D^{-1}(1/2) = \tilde{M}$ median

The Newsvendor Problem: normally distributed demand

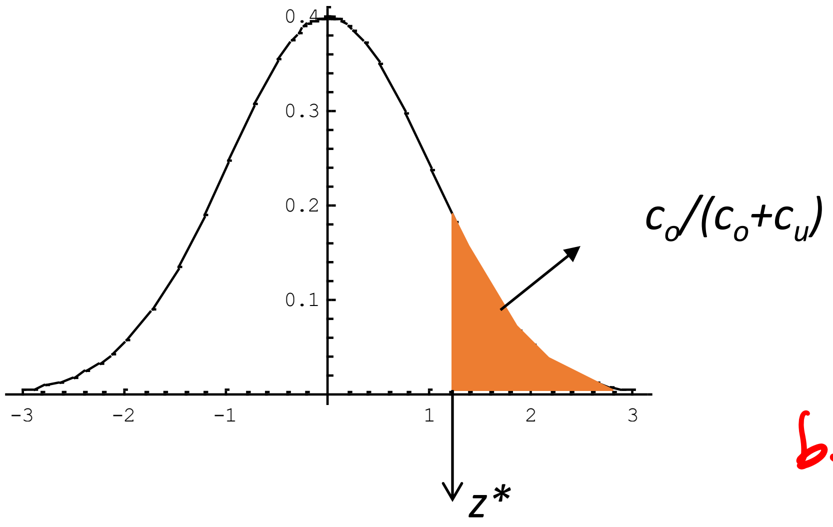
D : Normal (μ,σ)

$$F_D(Q^*) = \frac{c_u}{c_u + c_o}$$

Define $Z=(D-\mu)/\sigma$, we are looking for z^* such that:

$$F_Z(z^*) = \frac{c_u}{c_u + c_o}$$

The Newsvendor Problem: normally distributed demand



but if
$c_o > c_u$
$\Rightarrow z^* < 0$

$$\Rightarrow Q^* = \mu + z^* \sigma$$

Interpretation : z^* – safety factor (depends only on the financial parameters)

optimal order quantity = mean demand (μ)+ safety stock ($z^*\sigma$)

The Newsvendor Problem: normally distributed demand

Let f_z be the pdf of Z (normal (0,1)):

$$Q^* = \mu + z^*\sigma,$$

The optimal cost :

$$E[C(Q^*)] = E[c_u(D - Q^*)^+ + c_o(Q^* - D)^+] = (c_u + c_o)f_Z(z^*)\sigma$$

This is a simplification that only works for a normal distribution.
The optimal cost does not depend on the average demand !

Expected Profit

The expected optimal profit is:

$$E[\Pi(Q^*)] = (c_u)\mu \; - \; E[C(Q^*)]$$
$$= (c_u)\mu - E[c_u(D - Q^*)^+ + c_o(Q^* - D)^+]$$
$$= (c_u)\mu - \; (c_u + c_o)f(z^*)\sigma$$

Note that the expected profit depends on the average demand.
Moreover, it can be negative if:

$$(c_u)\mu < \; (c_u + c_o)f(z^*)\sigma$$

Prediction based on a regression

- Assume that the prediction of demand is based on a regression model:

$$d_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_p x_{pt} + \epsilon_t$$

- Assume that the most recent values of the predictors are: $x'_1, x'_2 ..., x'_p$. If the regression errors are normally distributed with standard deviation $\hat{\sigma}$ We then have:

$$Q^* = (\beta_0 + \beta_1 x'_1 + \beta_2 x'_2 + ... + \beta_p x'_p) + z^* \hat{\sigma}$$

- If we let $\beta_{00} = \beta_0 + z^* \hat{\sigma}$, we can note that optimal order quantity is linear in the predictors:

$$Q^* = \beta_{00} + \beta_1 x'_1 + \beta_2 x'_2 + ... + \beta_p x'_p$$

Prediction based on a regression

- We can also get a closed-form expression for the optimal profit:

$$E[\Pi(Q^*)] = c_u(\beta_0 + \beta_1 x_1' + \beta_2 x_2' + ... + \beta_p x_p') - (c_u + c_o) f_Z(z^*)\hat{\sigma}$$

- We can the see that there are some general statements to make:
  1. If $\beta_i > 0$, then when predictor x_i' increases, expected optimal profit increases
  2. If $\beta_i < 0$, then when predictor x_i' increases, expected optimal profit decreases
  3. Expected optimal profit always decreases in $\hat{\sigma}$

Discrete Demand

- When demand arrives in small discrete quantities, the continuous approximation is not reasonable
- There are many practically relevant discrete distributions
    - Poisson (discrete analogue to Normal distribution)
    - Negative binomial (known to be a good model of retail demand)
    - And of course, discrete empirical distributions (each value from a sample of n past demand observations has equal probability $1/n$).
- In a data-based approach, we might directly want to work with the (training) sample: $y_1, y_2, \ldots y_n$.

Discrete Optimization

$$E[C(Q\quad)] = E[c_u(D - Q\quad)^+ + c_o(Q\quad - D)^+]$$

- The expected cost (or profit) function is no longer differentiable
- But we can still show discrete convexity (concavity) of the objective function
- A discrete function $f(x)$ $(x \in Z)$ is convex if:

$$f(x+1) - f(x) \geq f(x) - f(x-1), \forall x \in Z$$

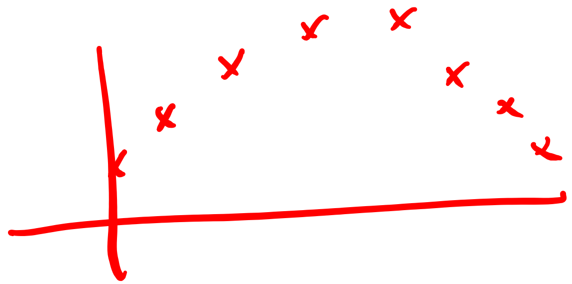
Discrete distributions:

marginal analysis

D has a discrete distribution.

$P(D=0) = p_D(0)$, $P(D=i) = p_D(i)$

$$E[\pi(Q+1)] - E[\pi(Q)] = -w + r\,P(D>Q) + s\,P(D \le Q)$$



$$E[\pi(Q+2)] - E[\pi(Q+1)] \overset{?}{\le} E[\pi(Q+1)] - E[\pi(Q)]$$

Discrete Optimization

- Let's take a look at the first difference of the expected cost as a function of Q.

$$E[C(Q+1) - C(Q)] = w - r\,P(D > Q) - sP(D \leq Q)$$
$$E[C(Q) - C(Q-1)] = w - r\,P(D > Q-1) - sP(D \leq Q-1)$$

- The second difference is:

$$-r\,\big(P(D > Q) - P(D > Q-1)\big) + (sP(D \leq Q)\text{-}\,sP(D \leq Q-1) =$$
$$rP(D = Q) - s\big(P(D = Q)\big) = (r-s)P(D = Q) \geq 0$$

- The expected cost function is therefore discrete convex

Discrete Optimization

- The expected cost function is discrete convex
  - Its first difference will be non-decreasing and can switch from negative to positive only once. The point where the first difference changes sign is the minimizer (it may not be unique).

- We then look for Q where the first difference, $E[C(Q+1) - C(Q)] = w - r\,P(D > Q) - sP(D \leq Q)$ changes sign. This gives:

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in \mathrm{Z}^+\right\}$$

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in Z^+\right\}$$

Ex: $P(D=0) = 0.5$, $P(D=1) = 0.3$, $P(D=2) = 0.2$

x	$F_D(x)$
0	0.5
1	0.8
2	1

let $\frac{c_u}{c_u + c_o} = 0.9$

$\Rightarrow Q^* = 2$

if $\frac{c_u}{c_u + c_o} = 0.4$

$Q^* = 0$

Ex 2: Poisson with mean 2

$P(D=i) = \frac{2^i e^{-2}}{i!}$ $\quad i = 0, 1, 2 \ldots$

$F_D(x) = \Sigma$

Discrete demand example

- Let D have a geometric distribution with parameter p

$$P\{D = x\} = (1-p)p^x \quad \text{for } x = 0,1,2,...$$

$$P\{D \le x\} = \sum_{i=0}^{x}(1-p)p^i = 1 - p^{x+1}$$

$$\Rightarrow Q^* = \left\lfloor \frac{\log\left(\frac{c_o}{c_u + c_o}\right)}{\log(p)} \right\rfloor$$

Discrete demand example

- Consider the following training observations (Toyota Sales).

Month	Sales
2006-01-01	808
2006-02-01	969
2006-03-01	1475
2006-04-01	1439
2006-05-01	1394
2006-06-01	1364
2006-07-01	890
2006-08-01	896
2006-09-01	1068
2006-10-01	968
2006-11-01	783
2006-12-01	1346

Discrete demand example

- Assume that we would like to set a safety stock target and $c_u/(c_u+c_o) = 0.8$.

Month	Index	Sales	Sorted	F(x)
2006-01-01	1	808	783	0.083
2006-02-01	2	969	808	0.167
2006-03-01	3	1475	890	0.250
2006-04-01	4	1439	896	0.333
2006-05-01	5	1394	968	0.417
2006-06-01	6	1364	969	0.500
2006-07-01	7	890	1068	0.583
2006-08-01	8	896	1346	0.667
2006-09-01	9	1068	1364	0.750
2006-10-01	10	968	1394	0.833
2006-11-01	11	783	1439	0.917
2006-12-01	12	1346	1475	1.000

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in \mathrm{Z}^+\right\}$$

$$Q^* = 1394$$

Empirical Risk Minimization

- What we did is to take the observed sample as the ‘world’, each observation was assumed to take place with probability 1/12.
- The downside is that we are limited with the small sample: if we observed a demand of 968 in the past maybe we’ll observe demands of 967 or 969 in the future but those are not part of our world.
- The upside is that we did not attempt to fit a distribution and therefore avoided the estimation errors.

Empirical Risk Minimization

- Empirical Risk Minimization is useful if we don't have a formula for the optimal decision such as the following one.

$$Q^* = \min\left\{F_D(Q) \geq \frac{c_u}{c_u + c_0},\ Q \in \mathrm{Z}^+\right\}$$

- We can handle other expected cost minimization formulations subject to constraints.
- Let us take the example of a constrained optimization formulation.

Empirical Risk Minimization

Here's a stochastic optimization formulation that is a linear program:

$$
\begin{aligned}
&\min_{Q} R(Q) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q)^+ + c_o(Q - d_i)^+ \\
&\equiv \min_{Q} \frac{1}{n}\sum_{i=1}^{n} c_u\, z_i{}^+ + c_o z_i{}^- \\
&s.t. \\
&\quad z_i{}^+ \quad \geq \quad d_i - Q \quad i = 1,2,\ldots,n \\
&\quad z_i{}^- \quad \geq \quad Q - d_i \quad i = 1,2,\ldots,n \\
&z_i{}^+, z_i{}^- \quad \geq \quad 0 \quad i = 1,2,\ldots,n
\end{aligned}
$$

Empirical Risk Minimization

- The stochastic optimization formulation has recently received a lot of attention.
- We'll next look at a more recent formulation that combines predictive analytics with prescriptive analytics.

Gah-Yi Ban and Cynthia Rudin, "The Big Data Newsvendor: Practical Insights From Machine Learning", *Operations Research*, Vol. 67, pp. 90-108, 2019.

Forecasting demand using predictive analytics

- Ice-cream store: Daily demand depends on
  - Day of week
  - Temperature
  - Weather condition (sunny, cloudy, rainy etc.)
- What is the demand for tomorrow?
  - Friday
  - Forecasted temperature 19^{o}C.
  - Partly sunny
- We have seen many approaches to handle such predictors (features): simple regression, non-linear regressions, random forests etc.

Forecasting demand

- Ice-cream store
- Naive model: ignore the dependence on the predictors.
- Then optimal order quantity for Friday (or any other day):

$$Q^{*}=F_{D}^{-1}(c_u/c_u+c_o)$$

where F_D is the empirical distribution coming from past demand data.

The separated optimization framework

- Ice-cream store: better model
- Model the dependence on the predictors:
- $D = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ where ε is Normal$(0, \sigma_\varepsilon)$.
- Estimate coefficients $\beta_0, \beta_1, \beta_2, \beta_3$ (by regression)
- Then, optimal order quantity for Friday:

$$Q^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + z^* \sigma_\varepsilon.$$

where (x_1, x_2, x_3) are the observed predictors.

Separated vs. Joint Optimization

- The above approach is standard we first use predictive analytics to estimate the demand and then solve an optimization problem for the optimization approach.
- We can call this separated estimation and optimization (estimation and optimization stages are clearly separated).
- Ban and Rudin (2019) take an alternative approach and propose a joint estimation and optimization approach.

Feature based newsvendor

- Ice-cream store
- Demand depends on observable features. Let $\boldsymbol{x}$ be the feature vector.
- Then the cost minimization problem is

$$\min_{Q(\boldsymbol{x})} E[C(Q(\boldsymbol{x}); D(\boldsymbol{x}) \mid \boldsymbol{x})]$$

Feature based newsvendor: data

- Now assume that past demand data as a function of the feature is available: we have observations: $S_n = (d_1,\mathbf{x}_1), (d_2,\mathbf{x}_2),\ldots, (d_n,\mathbf{x}_n))$.
- For the ice cream vendor:

Data Point	Day	Temp.	Weather	Sales
1	Monday	12	Rainy	23kg
2	Tuesday	14	Cloudy	28kg
3	Wed.	14	Sunny	35kg
4	Thu.	17	Sunny	30kg
…	…	….	…	…

Feature based newsvendor: machine learning ideas

- Formulate and find a way to solve the following problem:

$$\min_{Q(\cdot)} R(Q(\cdot), S_n) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q(\mathrm{x}_i))^+ + c_o(Q(\mathrm{x}_i) - d_i)^+$$

- To solve the above, we need to guess the functional form of $Q(\boldsymbol{x})$.
- Reasonable guess: a linear decision rule

$$Q(\mathrm{x}) = Q((x^1, x^2, \ldots, x^p)) = q^0 + \sum_{j=1}^{p} q^j\, x^j$$

- Note that this is supported by what we have seen before: the optimal order quantity is a linear function of the predictors and the standard deviation of the estimation error

Feature based newsvendor: machine learning ideas

- Here's the ML – optimization formulation:

$$\min_{Q(\cdot)} R(Q(\cdot), S_n) = \frac{1}{n}\sum_{i=1}^{n} c_u\,(d_i - Q(\mathrm{x}_i))^+ + c_o(Q(\mathrm{x}_i) - d_i)^+$$

$$\equiv \min_{Q=(q^1,q^2,\ldots q^p)} \frac{1}{n}\sum_{i=1}^{n} c_u\, z_i{}^+ + c_o z_i{}^-$$

$$s.t.$$

$$\begin{aligned}
z_i{}^+ \quad &\geq \quad d_i - \left(q^0 + \sum_{j=1}^{p} q^j x_i^j\right) && i = 1,2,,\ldots,n \\
z_i{}^- \quad &\geq \quad q^0 + \sum_{j=1}^{p} q^j x_i^j - d_i && i = 1,2,,\ldots,n \\
z_i{}^+, z_i{}^- \quad &\geq \quad 0 && i = 1,2,,\ldots,n
\end{aligned}$$