Pricing Related Goods

Jonathan Stokes

December 31, 2020

Be forewarned that while I think I have some pretty numerical results, the analytical side is still a bit rough. For instance I would love to be able to prove the concavity of the profit equations, but perhaps another day. Also this entire set of notes is a rough draft, for what I'm unsure. A Java implementation of the profit optimization algorithm can be found here `https://github.com/sophist0/opt_profit`

<u>Two Good Model</u>

Assume we have two goods labeled G_1 and G_2 where G_2 is a sequel to G_1. For instance G_1 is Home Alone and G_2 is Home Alone 2. These have a potential audience A. The fraction of the audience that wants to good G_1 first is $f_{0,1}$. The faction that wants to see movie M_2 first is $f_{0,2}$. The fraction of viewers that see G_1 first and then want to see G_2 is $f_{1,2}$ and the fraction of viewers that see G_2 first and then want see G_1 is $f_{2,1}$. These fractions are subject to the following restriction $0 \leq f_{i,j} \leq 1$ for $i \in [0,1,2]$ and $j \in [1,2]$. Additionally no sum of fractions leaving an audience node may be greater than 1 which in this case implies that

$$0 \leq f_{0,1} + f_{0,2} \leq 1 \tag{1}$$

These fractions $f_{i,j}$ are often empirically available to online marketplaces such as YouTube Movies.

The possible orders in which the audience can purchase G_1 and G_2 can be represented as a graph.
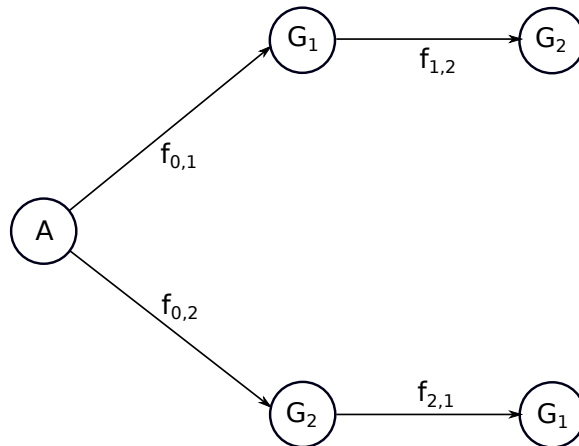

Figure 1: Two Good Purchase Order Model

To this model we can add the price p_i of each good. But before doing so lets normalize the min and max pricing of the goods such that at $p_i = 0$ everyone who wants good G_i purchases it and at $p_i = 1$ no one who wants good G_i purchases it. As with the fractions $f_{i,j}$ the actual min and max prices of a good can be estimated empirically. At this point I need to make and important clarification. I assume the min and max prices of all goods are the same. This makes sense in the case of two movies, but is harder to justify

if the goods in question were a pound of coffee and a pound of sugar.

Define p_i as having domain $p_i \in [0, 1]$. Finally I assume that every member of the audience is subject to the same linear price sensitivity function $s_i = (1 - p_i)$ so at $p_i = 1/2$ half the members of the audience who want to purchase good G_i do so. I have no reason to think that this price sensitivity function reflects actual consumer behavior, I chose it for its simplicity. The resulting two good pricing model is given below,
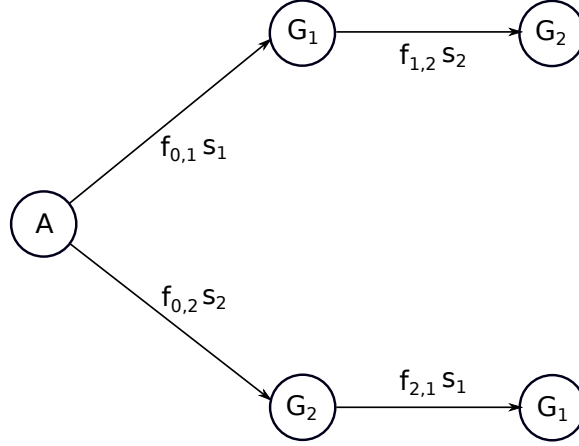

Figure 2: Two Good Pricing Model

If we do not assume that all goods have the same min and max price I think the sensitivity function could be formulated as

$$s_i = \frac{p_i^+ - p_i}{p_i^+ - p_i^-} \tag{2}$$

where p_i^+ is the max price of good i and p_i^- the min price. But in truth I have not entirely worked though this case yet.

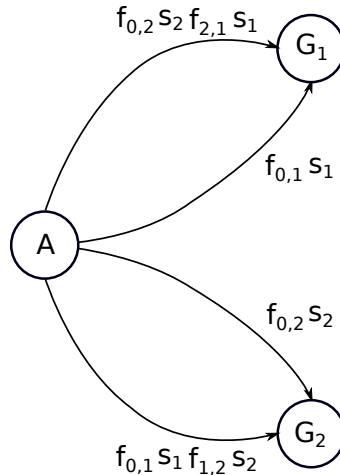The model in the figure above can be collapsed to the following graph,

Figure 3: Collapsed Two Good Pricing Model

Two Good Profit Equation

If the audience size is N its clear from the model above that the profit earned by both goods can be

2

computed as

$$profit := f(p_1, p2) \tag{3}$$
$$= N[(f_{0,1}s_1 + f_{0,2}s_2f_{2,1}s_1)p_1 + (f_{0,2}s_2 + f_{0,1}s_1f_{1,2}s_2)p_2] \tag{4}$$
$$\propto (f_{0,1}s_1 + f_{0,2}s_2f_{2,1}s_1)p_1 + (f_{0,2}s_2 + f_{0,1}s_1f_{1,2}s_2)p_2 \tag{5}$$
$$= [(f_{0,1}(1-p_1) + f_{0,2}f_{2,1}(1-p_1)(1-p_2)]p_1 + [(f_{0,2}(1-p_2) + f_{0,1}f_{1,3}(1-p_1)(1-p_2))]p_2 \tag{6}$$
$$= (p_1 - p_1^2)(f_{0,1} + f_{0,2}f_{2,1}(1-p_2)) + (p_2 - p_2^2)(f_{0,2} + f_{0,1}f_{1,2}(1-p_1)) \tag{7}$$

Lets simplify the notation using the following mapping.

$$p_1 \to x, \ p_2 \to y, \ f_{0,1} \to a, \ f_{0,2} \to b, \ f_{0,1}f_{1,2} \to c, \ f_{0,2}f_{2,1} \to d \tag{8}$$

Under this mapping the equation simplifies to

$$g(x,y) = \frac{f(x,y)}{N} = (x - x^2)(a + d(1-y)) + (y - y^2)(b + c(1-x)) \tag{9}$$

Taking the first and second partial derivates of $g(x,y)$ gives

$$\frac{\partial}{\partial x}g(x,y) = (1 - 2x)(a + d(1-y)) - c(y - y^2) \tag{10}$$

$$\frac{\partial}{\partial y}g(x,y) = (1 - 2y)(b + c(1-x)) - d(x - x^2) \tag{11}$$

$$\frac{\partial^2}{\partial x^2}g(x,y) = -2(a + d(1-y)) \tag{12}$$

$$\frac{\partial^2}{\partial y^2}g(x,y) = -2(a + d(1-x)) \tag{13}$$

$$\frac{\partial^2}{\partial x \partial y}g(x,y) = 2d\left(x - \frac{1}{2}\right) - 2c\left(y - \frac{1}{2}\right) \tag{14}$$

$$\frac{\partial^2}{\partial y \partial x}g(x,y) = 2d\left(x - \frac{1}{2}\right) - 2c\left(y - \frac{1}{2}\right) \tag{15}$$

Therefore the Hessian matrix of $g(x,y)$ is

$$Hg(x,y) = \begin{bmatrix} \frac{\partial^2}{\partial x^2} & \frac{\partial^2}{\partial x \partial y} \\ \frac{\partial^2}{\partial y \partial x} & \frac{\partial^2}{\partial y^2} \end{bmatrix} = \begin{bmatrix} -2[a + d(1-y)] & 2[d(x - 1/2) - c(y - 1/2)] \\ 2[d(x - 1/2) - c(y - 1/2)] & -2[b + c(1-x)] \end{bmatrix} \tag{16}$$

Results for the Two Good Profit Equation

The squared terms in the derivative of $g(x,y)$ make finding a closed form solution for the critical points of $g(x,y)$ difficult in terms of arbitrary parameterizations of $f_{0,1}$, $f_{0,2}$, $f_{1,2}$, and $f_{2,1}$. Therefore before solving for these points numerically I set the values of these parameters as follows,

$$f_{0,1} = \frac{9}{10}, \ f_{0,2} = \frac{1}{10}, \ f_{1,2} = \frac{3}{4}, \ f_{2,1} = \frac{1}{4} \tag{17}$$

Practically these fractions could be found empirically by setting $p_1 = 0$ and $p_2 = 0$. Using these values under the mapping defined above gives

$$a = \frac{9}{10}, \ b = \frac{1}{10}, \ c = \frac{27}{40}, \ d = \frac{1}{40} \tag{18}$$

Plugging these values into the partial derivatives of $g(x,y)$ gives

$$\frac{\partial}{\partial x}g(x,y) = (1 - 2x)\left(\frac{9}{10} + \frac{1}{40}(1-y)\right) - \frac{27}{40}(y - y^2) = 0 \tag{19}$$

$$\frac{\partial}{\partial y}g(x,y) = (1 - 2y)\left(\frac{1}{10} + \frac{27}{40}(1-x)\right) - \frac{1}{40}(x - x^2) = 0 \tag{20}$$

3

Solving this system of equations with Sympy's nsolve and initial points $x, y = 1/2$, I found the critical point for $g(x, y)$ is $x^* = 0.408$ and $y^* = 0.494$ rounding to the thousandths place.

Plugging the values into the Hessian matrix we can determine if this solution a saddle point or a local maxima using the partial derivative test.

$$Hg(x^*, y^*) = \begin{bmatrix} -1.8253 & 0.0035 \\ 0.0035 & -0.9992 \end{bmatrix} \tag{21}$$

The computing eigenvalues of $Hg(x^*, y^*)$ with numpy gives $\lambda_1 = -1.8253$ and $\lambda_2 = -0.9992$. Since all the eigenvalues of the Hessian are negative, the Hessian is a negative definite matrix. Therefore $g(x, y)$ achieves a local maxima at point $(0.408, 0.494)$.

Is this solution better than if we fixed $p_1 = p_2$ in the Two Good Pricing Model above? If we set $x = y$, $g(x, y)$ reduces to

$$g(x) = x^3(c + d) - x^2(a + b + 2(c + d)) + x(a + b + c + d) \tag{22}$$

and the derivatives of $g(x)$ are

$$\frac{d}{dx}g(x) = 3x^2(c + d) - 2x(a + b + 2(c + d)) + (a + b + c + d) \tag{23}$$

$$\frac{d^2}{dx^2}g(x) = 6x(c + d) - 2(a + b + 2(c + d)) \tag{24}$$

Setting the first derivative of $g(x)$ to zero and solving for the two critical points gives $x^* \in 0.438, 1.848$. Since by definition $0 \leq x < 1$ it follows that $x^* = 0.438$.
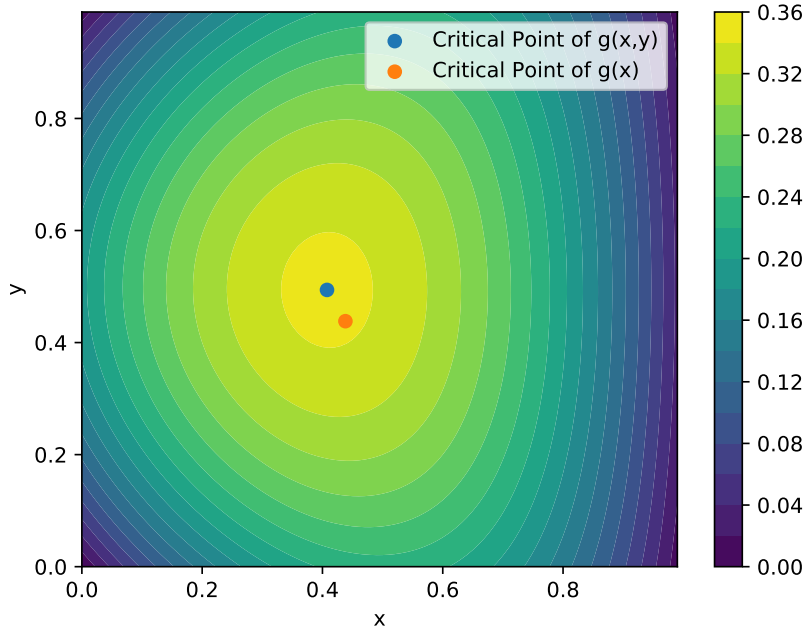

Figure 4: Contour Plot of $g(x, y)$

Now looking at the second derivative of $g(x)$ at x^* we find that

$$\frac{d^2}{dx^2}g(x^*) = -0.560 < 0 \tag{25}$$

Therefore $g(x^*)$ is a local maxima by the second derivative test and the only maxima in the domain of x.

What is the difference in profit achieved using $g(x, y)$ vs $g(x)$ to optimize pricing? Well

$$g(x^*) = 0.34299 \tag{26}$$

$$g(x^*, y^*) = 0.34532 \tag{27}$$

$$\frac{g(x^*, y^*)}{g(x^*)} = 1.00678 \tag{28}$$

or about a 0.68% increase as a result of price optimizing via $g(x, y)$. But recall this is the normalized profit per potential audience member. If the maximum price anyone is willing to pay for G_1 and G_2 is $p_{max} = \$20$ while the minimum is $p_{min} = \$0$, this sets $p_1 = \$8.16$, $p_2 = \$9.88$ under $g(x, y)$ and $p_1 = p_2 = \$8.76$ under $g(x)$. Now if the audience size is 10^8 the difference in profit using $g(x, y)$ to set prices vs $g(x)$ is $\$13,560,000$ which is not trivial.

<u>n Good Model</u>

From the small example given in Fig. 2 its clear that to compute the profit over a set of goods purchased in sequence requires looking at every permutation of that sequence. The figure below gives a tree representing these permutations for a sequence of n goods.
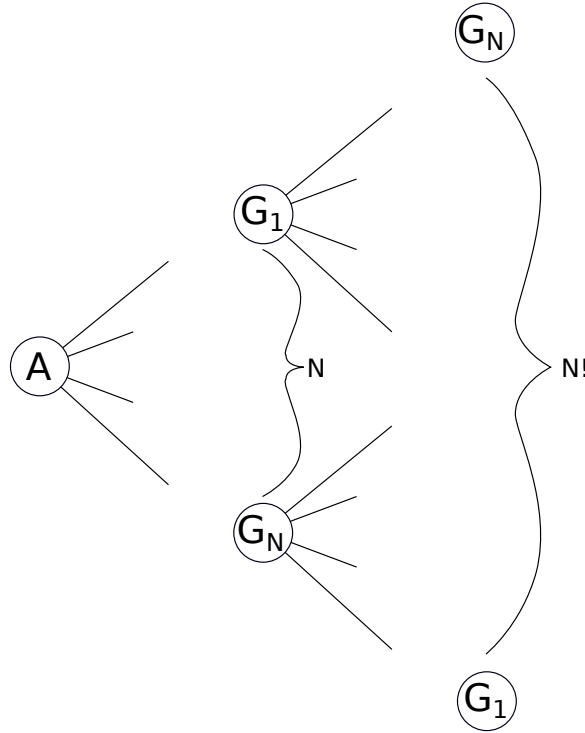
Figure 5: n Good Model

This tree has the following statistics.

- The n goods have $n!$ permutations and therefore this tree has $n!$ leaves.

- The number of edges in the tree can be computed as

$$edges = n + n(n-1) + n(n-1)(n-2) + \ldots \tag{29}$$

$$= \sum_{i=0}^{n-1} \left(\prod_{j=0}^{i} (n-j) \right) \tag{30}$$

- There are $2\binom{n}{2} + n$ unique parameters representing fractions $f_{i,j}$ where the final term captures the fraction of the audience that wants to purchase each good first i.e. fractions $f_{0,j}$.

If the graph of the n Good Model is collapsed as we collapsed the graph of the Two Good Model the result is n graphs of the form below each with $n!$ weighted edges from the audience to a good G_l.
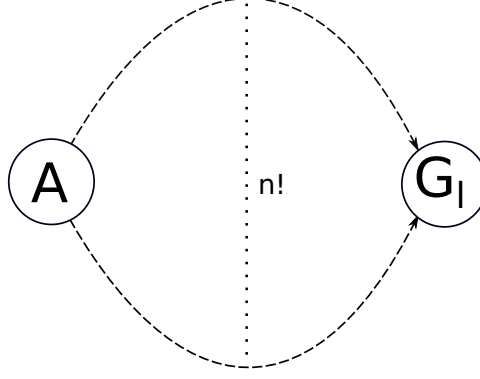

Figure 6: Collapsed n Goods Model

Each of these $n!$ edges is weighted by the product of between 1 and n terms $f_{x,y}(1 - p_j)$ depending on where in the permutation G_l is and each of these weighted edges is multiplied by p_l the price of good G_l giving equation for calculating the profit of good G_l along the k_{th} edge the form,

$$g_{l,k}(p_j, \ldots, p_k) = f_{x,y}(1 - p_j) \cdot f_{q,r}(1 - p_l) \cdots f_{w,z}(1 - p_k)p_l \tag{31}$$

where the total profit is

$$profit = f(p_1, \ldots, p_n) = \sum_{l=1}^{n} g_l(p_1, \ldots, p_n) = \sum_{l=1}^{n} \sum_{k=1}^{n!} g_{l,k}(p_j, \ldots, p_k) \tag{32}$$

Taking the second partial derivatives of $g_{l,k}(p_1, \ldots, p_n)$ gives

$$\frac{\partial^2}{\partial p_l^2} g_{l,k}(p_j, \ldots, p_k) = -2f_{x,y}(1 - p_j) \cdots f_{u,v}(1 - p_{l-1})f_{q,r}f_{e,o}(1 - p_{l+1}) \cdots f_{w,z}(1 - p_k) \tag{33}$$

$$\frac{\partial^2}{\partial p_i \partial p_j} g_l(p_j, \ldots, p_k) = 0 \qquad \text{for } i \neq l \tag{34}$$

$$\frac{\partial^2}{\partial p_l \partial p_j} g_{l,k}(p_j, \ldots, p_k) = -f_{x,y}f_{q,r}(1 - 2p_l) \cdots f_{w,z}(1 - p_k) \tag{35}$$

To simplify notation lets define this second partial derivatives of $g_l(p_j, \ldots, p_k)$ and $g_{l,k}(p_j, \ldots, p_k)$ as

$$g_{l,k}''(l, j) := \frac{\partial^2}{\partial p_l \partial p_j} g_{l,k}(p_j, \ldots, p_k) \tag{36}$$

$$g_l''(l, j) := \frac{\partial^2}{\partial p_l \partial p_j} g_l(p_j, \ldots, p_k) \tag{37}$$

where $l, j \in [1, \ldots, n]$. It follows from the sum rule for derivatives that

$$g_l''(l, j) = \sum_{k=1}^{n!} g_{l,k}''(l, j) \tag{38}$$

Notice that the functions $g_l''(l, j)$ are the l, j entries of the Hessian matrix for the profit function $g(p_1, \ldots, p_n)$ such that

$$Hg(p_1, \ldots, p_n) = \begin{bmatrix} g_1''(1, 1) & g_1''(1, 2) & \cdots & g_1''(1, n) \\ g_2''(2, 1) & g_2''(2, 2) & \cdots & g_2''(2, n) \\ \vdots & \vdots & \cdots & \vdots \\ g_n''(n, 1) & g_n''(n, 2) & \cdots & g_n''(n, n) \end{bmatrix} \tag{39}$$

6

Which in theory could be computed for given a set of prices and audience fraction parameters for an arbitrary n. This Hessian matrix could then be used to determine if the price point is a local maxima.

The Profit Algorithm

In terms of computing profit over all permutations of a length n sequence, the graph representation is irrelevant. The profit can be directly computed from the permutations. The algorithm I give below while inefficient is straight forward.

If $\mathcal{G}$ is the set of goods with arbitrary initial ordering $\boldsymbol{\sigma}_0 \equiv [1, \ldots, n]$ where $n = |\mathcal{G}|$, and $\mathbf{p} = [0, p_1, \ldots, p_n]$ is the prices of these goods with the 0 added to simplify indexing. The matrix $\mathbf{F} = [i, j] \in \mathbf{M}_{n+1}$ has entries which if $i = 0$ is the fraction of individuals in the audience or if $i > 0$ the faction of the audience who have purchased goods $\boldsymbol{\sigma}_{x,1}$ to $\boldsymbol{\sigma}_{x,i}$ and would purchase good $\boldsymbol{\sigma}_{x,j}$ if $p_j = 0$ where $\boldsymbol{\sigma}_{x,j}$ is the j_{th} good in the x_{th} permutation.

The set of permutations of the goods in set $\mathcal{G}$ is denoted $\mathcal{P}_n$ where $|\mathcal{P}_n| = n!$. Additionally if $a \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^l$ let $\oplus$ denote the concatenation of an element with a vector such that $a \oplus \mathbf{v} = \mathbf{w}$ where $\mathbf{w}_0 = a$ and $\mathbf{w} \in \mathbb{R}^{l+1}$. Finally let $|\mathbf{w}|$ denote the length of $\mathbf{w}$. Below is the algorithm used to compute the profit.

$profit = 0$
for $\boldsymbol{\sigma}_i \in \mathcal{P}_n$ **do**
    $c = 1$ ▷ Current fraction of the audience
    $b = 0$ ▷ Profit over the current permutation
    $\boldsymbol{\pi} = 0 \oplus \boldsymbol{\sigma}_i$
    for $j \in [0, \ldots, |\boldsymbol{\pi}| - 1]$ **do**
        $l = \boldsymbol{\pi}_j$
        $m = \boldsymbol{\pi}_{j+1}$
        $c = c(1 - \mathbf{p}_m)\mathbf{F}_{l,m}$
        $b = b + c\mathbf{p}_m$
    end for
    $profit = profit + b$
end for

Using stochastic gradient ascent (SGA) to optimize the Two Good Model with the parameters f_1, f_2, f_3, f_4 given and computing the gradient of profit using the algorithm above as a function of the prices $\mathbf{p}$. The SGA algorithm initiates all prices to $1/2$, the learning rate is $\nu = 0.01$, and it iterates 1000 times, resulting in $profit = 0.34529$ for $x = 0.403$, $y = 0.489$ which is the same to the thousandths place as the local maxima of $g(x, y)$ found above.

Results for the n Good Profit Equation

The SGA optimization algorithm can be used to optimize larger sets of goods and is limited only by the difficulty of computing the profit over the $n!$ permutations of the goods. This difficulty can be mitigated if its assumed most these permutations do not occur. I think the number of permutations that are likely to occur can be quantified using typical sets. But I have not done any pruning of the permutation space here. If we assume the fractions of the audience who would like to purchase the n goods are defined as

follows.

$$f_{0,i} = 2^{-i} \tag{40}$$

$$f_{i,i+1} = \frac{9}{10} \text{ for } i > 0 \tag{41}$$

$$f_{i,j} = \frac{1}{100} \text{ for } j \neq i + 1 \tag{42}$$

We can look at how the optimal prices vary with n as computed by the SGA given $\nu = 0.005$ in the table below,

n	p_1^*	p_2^*	p_3^*	p_4^*	p_5^*	p_6^*	$g(p_1, \ldots p_n)$	$g(p)$	$g(p_1, \ldots p_n)/g(p)$
1	0.495						0.1250	0.1250	1.000000
2	0.384	0.494					0.2504	0.2472	1.012944
3	0.411	0.419	0.494				0.5598	0.5572	1.004666
4	0.443	0.438	0.444	0.494			1.6751	1.6736	1.000896
5	0.460	0.457	0.456	0.460	0.494		6.6159	6.6145	1.000211
6	0.468	0.467	0.466	0.466	0.469	0.494	32.7161	32.7143	1.000055

It is interesting that the last price is relatively stable and proceeding prices are nearly symmetric. It is also interesting that the ratio of profit under independent pricing and identical pricing decreases with n increasing. Both of these phenomina are likely due to the parameters $f_{i,j}$ I have choosen rather than anything inherent to the structure of optimally pricing related goods. But to have any confidence in this statement would require additional exploration.

Puring Permutations

Consider the following tree which generates all permuations of $\{1, 2, 3\}$ prepended with a 0. Lets call this tree the source **X**. Let X_i be the i_{th} permutation generated by **X** and Y_j the j_{th} value in a permutation sometimes specified as $\mathsf{Y}_{i,j}$ if it is required to indicate that its the j_{th} value of the i_{th} permutation. The fractions in this tree were chosen independently at random according to the following distributions.

$$Pr(\mathsf{f}_{0,j} = \frac{1}{3}) = 1, \qquad Pr(\mathsf{f}_{0,j} = \frac{1}{6}) = 0; \qquad Pr(\mathsf{f}_{i,j} = \frac{1}{3}) = \frac{1}{3}, \qquad Pr(\mathsf{f}_{i,j} = \frac{1}{6}) = \frac{2}{3} \tag{43}$$
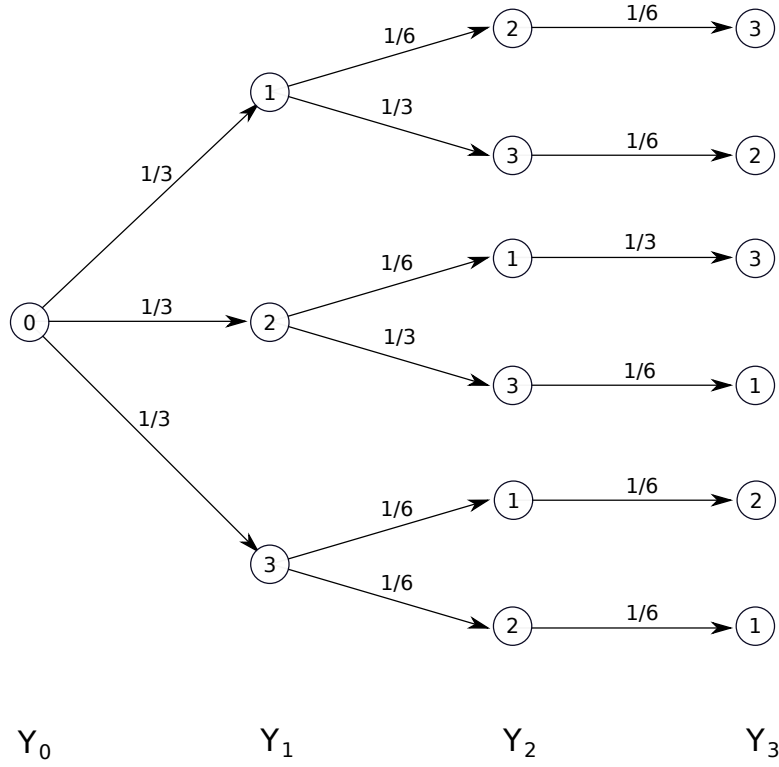
Figure 7: Source $\mathbf{X}$

The probabilities of Y_j taking on a value in a message X_i from $\mathbf{X}$ is given in the table below.

Y_0	Y_1	Y_2	Y_3
$P(\mathsf{Y}_0 = 0) = 1$	$P(\mathsf{Y}_1 = 0) = 0$	$P(\mathsf{Y}_2 = 0) = 0$	$P(\mathsf{Y}_3 = 0) = 0$
$P(\mathsf{Y}_0 = 1) = 0$	$P(\mathsf{Y}_1 = 1) = \frac{1}{3}$	$P(\mathsf{Y}_2 = 1) = \frac{2}{9}$	$P(\mathsf{Y}_3 = 1) = \frac{1}{36}$
$P(\mathsf{Y}_0 = 2) = 0$	$P(\mathsf{Y}_1 = 2) = \frac{1}{3}$	$P(\mathsf{Y}_2 = 2) = \frac{1}{9}$	$P(\mathsf{Y}_3 = 2) = \frac{1}{36}$
$P(\mathsf{Y}_0 = 3) = 0$	$P(\mathsf{Y}_1 = 3) = \frac{1}{3}$	$P(\mathsf{Y}_2 = 3) = \frac{2}{9}$	$P(\mathsf{Y}_3 = 3) = \frac{1}{27}$
$P(\mathsf{Y}_0 = \emptyset) = 0$	$P(\mathsf{Y}_1 = \emptyset) = 0$	$P(\mathsf{Y}_2 = \emptyset) = \frac{4}{9}$	$P(\mathsf{Y}_3 = \emptyset) = \frac{49}{54}$

Examining the table above its clear that the distribution of Y_j is not equal to the distribution of Y_{j-1}, therefore Y_j is not a stationary stochastic process. Additionally examining the source notice that $P(\mathsf{Y}_3 = 3 | \mathsf{Y}_2 = 2)$ may equal $\frac{1}{6}$ or 0, therefore Y_j is not a Markov process.

However if we consider the source $\mathbf{X}$ producing messages X_i the situation is very different. Each message X_i is independent and identically distributed according to distributions Y_j for $j \in [0, \ldots, n]$. Denoting the distribution of X_i explicity as F_{X_i} its also clear that the source $\mathbf{X}$ is strictly stationary since,

$$F_{\mathsf{X}_i} = F_{\mathsf{X}_{i+\tau}} \qquad \text{for all} \qquad i, \tau \in \mathbb{N} \tag{44}$$

Before continuing its necessary to define the convergence of a sequence of random variables $\mathsf{X}_1, \mathsf{X}_2, \ldots$ to a random variable X in probability if for every $\epsilon > 0$, $Pr\{|\mathsf{X}_n - \mathsf{X}| > \epsilon\} \to 0$. Given the notion of convergence in probablity the Asymptotic Equipartition Property (AEP) Theorem (Cover Thm. 3.1.1) can be stated as

$$-\frac{1}{n} \log p(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n) \to^{prob.} H(\mathsf{X}) \tag{45}$$

where $H(\mathsf{X})$ is the entropy of X.

Cover also defines a *typical set* $A_\epsilon^{(n)}$ for a sequence $(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n) \in \mathcal{X}^n$ with the property

$$2^{-n(H(\mathsf{X})+\epsilon)} \leq Pr(\mathsf{X}_1, \mathsf{X}_2, \ldots, \mathsf{X}_n) \leq 2^{-n(H(\mathsf{X})-\epsilon)} \tag{46}$$

In Thm 3.1.2 Cover states that the typical set $A_\epsilon^{(n)}$ has among others the following properties

$$Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon \qquad \text{for } n \text{ sufficiently large.} \tag{47}$$

$$|A_\epsilon^{(n)}| \leq 2^{n(H(\mathbf{X})+\epsilon)} \tag{48}$$

$$|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{n(H(\mathbf{X})-\epsilon)} \qquad \text{for } n \text{ sufficiently large.} \tag{49}$$

Given our definition of the source $\mathbf{X}$ and the messages it produces satisfying the AEP property it follows that if $\mathbf{X}$ produces sequences of length m the number of these typical sequences is less than $2^{m(H(\mathbf{X})+\epsilon)}$.

Now unformtunaty this does not tell us the number of unique permutations X_i in the length m sequences that make up the typical set $A_\epsilon^{(m)}$. Clearly its $\leq m2^{m(H(\mathbf{X})+\epsilon)}$ but its not clear this is a tight enouph bound to be useful. Especially given that there is no necessary relation between n the number of elements in the a permuation X_i and m the length of permutation sequences. So perhaps we are at a dead end here and may need to take a different approach.

First notice that the random variables Y_j for the j_{th} element in a permutation follow a catigorical distirbution and thus if one has access to that distribution permuations can be sampled directly. This while interesting is not useful if finding the parameters of the catigorical dist. requies generating every permutation? Probably not.

Say instead we set a threshold and say we want to compute the profit on any branch upto and including those with at least $\alpha = \frac{1}{54}$ probability. How does this effect the source $\mathbf{X}$
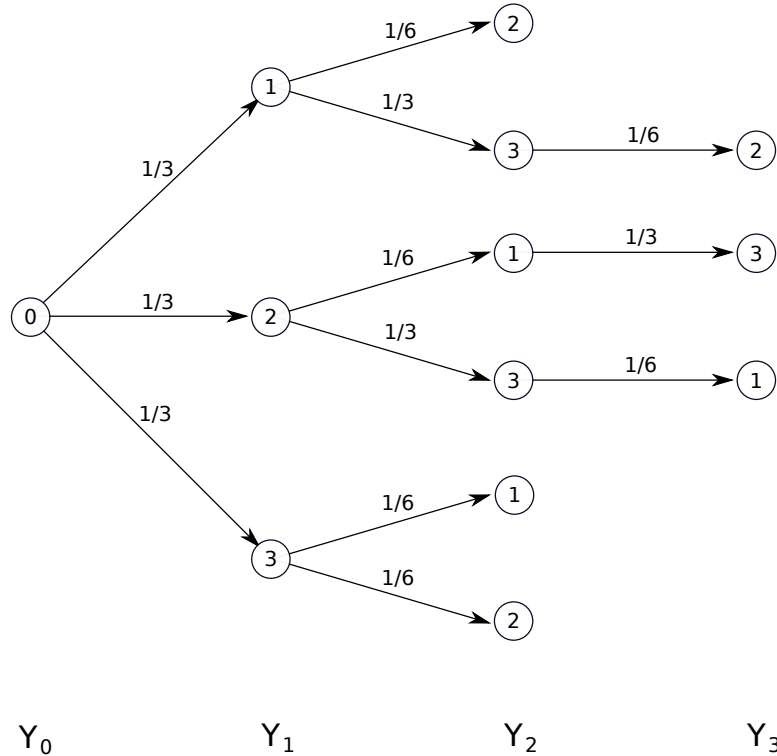

Figure 8: Truncated Source $\mathbf{X}$

Unfortunatly picking such a large α results in little pruing of such a small tree. Nor does setting thresold α give us how confident we should be that we captured $\beta\%$ of permutations in way that two standard deviations is guarenteed by the Chebyshev inequality to capture 75% of observations. However if we sum the probability at each root of the subtrees pruned from the truncated source we have the probability any message in the original source cannot be reproduced on the truncated source. To get at this quantity lets call $\mathcal{V}$ the set of node in the source tree $\mathbf{X}$ and v a node in this set. Additionally lets call $\mathbf{m}$ a message

10

from source **X** where **m** is a vector of nodes $v \in \mathcal{V}$ and denote the roots of the subtrees in the source **X** but not in the truncated source $\hat{\mathbf{X}}$ as set $\mathcal{R}$. $\mathcal{R}$ is the set of pruned subtree roots.

$$\beta = \sum_{v \in \mathcal{V}} \mathbb{1} \{v \in \mathcal{R}\} Pr(v \in \mathbf{m}) \tag{50}$$

The probability that a message **m** produced by source **X** cannot the reproduced by $\hat{\mathbf{X}}$ is equal to β since β is the probability that a message **m** terminates at a node in **X** but not in $\hat{\mathbf{X}}$. In the tree above $\beta = \frac{1}{36}$ so the truncated source $\hat{\mathbf{X}}$ correctly reproduces $\frac{35}{36}$ of the messages sent by source **X**. (Note this is different from the fraction of the profit captured by $\hat{\mathbf{X}}$ verse **X**.)

To construct a truncated source with a target β_t set α large and iteratively decrease α by ϵ until the difference between β of the truncated source and the target is acceptably small, $|\beta - \beta_t| < \delta$.

Next lets try to get a handle on the number of truncated permutations that exist for a given α without having to enumerate them. Let $\mathbf{f} = [f_v]$ be a vector containing the sum of the fraction weights leaving any node. So $f_r = 3(\frac{1}{3}) = 1$ the this sum for the root node. Define the following indicator function as

$$g(i, \alpha, \mathbf{f}) = \mathbb{1} \left\{ \prod_{j=1}^{i} \mathbf{f}_j \geq \alpha \right\} \tag{51}$$

and

$$c(\alpha, \mathbf{f}) = \sum_{i=1}^{n} g(i, \alpha, \mathbf{f}) \tag{52}$$

Also lets use $\mathbf{f}^-$ to denote that the vectors elements are in decreasing order and $\mathbf{f}^+$ to denote that its elements are in increasing order.

An upper bound on the number of permutations in the truncated tree can be given as

$$Permutations \leq \prod_{k=1}^{c(\alpha, \mathbf{f}^-)} k \tag{53}$$

And a lower bound on the number of permutations in the truncated tree can be given as

$$Permutations \geq \prod_{k=1}^{c(\alpha, \mathbf{f}^+)} k \tag{54}$$

The exepected number or permutations takes additional work. Instead of considering the actual edge weights $f_{i,j}$ consider the random variables $\mathbf{f}_{i,j}$ represnting the probabilities of those weights taking on specific values in our construction of the source. Since each permutation is prepended with a zero node representing the pool of people who may buy a good, $\mathbb{E}[\mathbf{f}_{0,j}]$ is likey different from $\mathbb{E}[\mathbf{f}_{i,j}]$ and in fact does so in our constructions. So we will define the two expecations seperately as follows

$$\bar{f}_0 = \mathbb{E}[\mathbf{f}_{0,j}] = \sum_f f Pr(\sum_j \mathbf{f}_{0,j} = f) \tag{55}$$

$$\bar{f}_{1+} = \mathbb{E}[\mathbf{f}_{i,j}] = \sum_f f Pr(\sum_{i,i \neq j} \sum_j \mathbf{f}_{i,j} = f) \tag{56}$$

(FIX Notation)and letting $\bar{\mathbf{f}} = [\bar{f}_0, \bar{f}_{1+}]$ the intermediate functions can be defined as

$$g(i, \alpha, \bar{\mathbf{f}}) = \mathbb{1} \left\{ \bar{\mathbf{f}}_0 (\bar{\mathbf{f}}_1)^{i-1} \geq \alpha \right\} \tag{57}$$

and

$$c(\alpha, \bar{\mathbf{f}}) = \sum_{i=1}^{n-1} g(i, \alpha, \bar{\mathbf{f}}) \tag{58}$$

Therefore it follows that

$$\mathbb{E}[Permutations] = \prod_{k=1}^{c(\alpha, \bar{\mathbf{f}})} k \tag{59}$$

Now there is a caveate with this expectation. Equation (57) can be a reduced to a product of expectations because we chose the values of the random variables $\mathsf{f}_{0,j}$ and $\mathsf{f}_{i,j}$ independently. This is not generally the case for sources and if it is not the case equation (57) is only an approximation of the expectation and an unbounded one at that. So the best one can do is upper and lower bound the number of permutations in the truncated source.

(I Need to test the bounds and the expectation!!!)

<u>References</u>

Elements of Information Theory 2ed. Thomas Cover.