DMA Modeling notes

Douglas Rubin

December 5, 2018

1 A Generative Model for Weighted Minutes Viewed

It is clear from the distributions of weighted minutes viewed per respondent per telecast (in each DMA) that a mixture model is appropriate. Specifically, there are 3 classes: respondents who do not watch the telecast and contribute zero weighted minutes, respondents who watch and contribute weighted minutes viewed with an exponential distribution, and respondents who watch the entire telecast. For the latter class, due to the weighting of the respondents, a Gaussian distribution, centered about the telecast show length is appropriate. In fact, this behavior is incredibly robust across tv program and DMA, and I show a few examples (with the mixture model fits) in Fig. 1.

Thus, to compute the total weighted minutes viewed for a specific telecast in a specific DMA, I first draw N_{DMA}, the number of respondents in the DMA at the time of the telecast:

$$N_{DMA} \sim \mathcal{D}(\Theta). \tag{1}$$

Now for $i = 1, 2, \ldots, N_{DMA}$, the rest of the generative model is:

$$Z^{(i)} \sim Cat(\phi_1, \phi_2, \phi_3) \tag{2}$$

$$\begin{cases} W^{(i)}|Z^{(i)} = 1 \sim \mathbb{1}\{w = 0\} \\ W^{(i)}|Z^{(i)} = 2 \sim Exp(\lambda) \\ W^{(i)}|Z^{(i)} = 3 \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \tag{3}$$

$$W_{tot}^{DMA} = \sum_{i=1}^{N_{DMA}} W^{(i)}. \tag{4}$$

Here $Z^{(i)}$ is a latent random variable which is the class of the i^{th} respondent $z^{(i)} \in \{1, 2, 3\}$, $W^{(i)}$ is the weighted minutes viewed for the i^{th} respondent and W_{tot}^{DMA}, which is total weighted minutes viewed for the telecast. Note that this is simply a mixture model whose parameters I can fit with maximum likelihood using the EM algorithm. Also note that technically a Gaussian truncated at zero would be more appropriate since $w^{(i)} \geq 0$, however this is not hugely problematic since for each DMA the mean of the Gaussian is quite far to the right of zero, so that virtually none of probability mass is less than zero.

The various PDFs/PMFs in the above generative model are

$$P(Z^{(i)} = z^{(i)}; \phi_1, \phi_2, \phi_3) = \phi_1^{\mathbb{1}\{z^{(i)}=1\}} \phi_2^{\mathbb{1}\{z^{(i)}=2\}} \phi_3^{\mathbb{1}\{z^{(i)}=3\}}, \tag{5}$$

and

$$\begin{cases} P(W^{(i)} = w^{(i)}|Z^{(i)} = 1) = \mathbb{1}\{w^{(i)} = 0\} \\ P(W^{(i)} = w^{(i)}|Z^{(i)} = 2; \lambda) = \lambda e^{-\lambda w^{(i)}} \\ P(W^{(i)} = w^{(i)}|Z^{(i)} = 3; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(w^{(i)}-\mu)^2}{\sigma^2}\right] \end{cases} \tag{6}$$

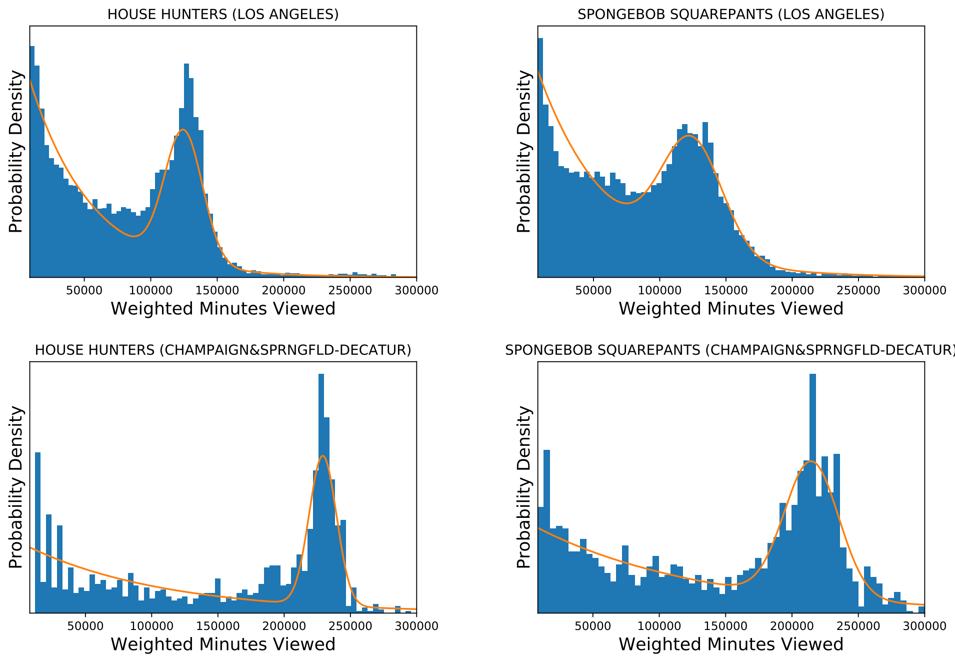


Figure 1: Distributions in weighted minutes viewed for 2 shows and 2 DMAs along with the corresponding mixture model fit. Note that the first class (that of weighted minutes viewed equals zero) is not included in these figures or the fits.

from which I can easily derive the PDF of $W^{(i)}$:

$$
\begin{aligned}
P(w^{(i)}) &= \sum_{z^{(i)}=1}^{3} P(w^{(i)}, z^{(i)}) && (7)\\
&= \sum_{z^{(i)}=1}^{3} P(w^{(i)}|z^{(i)})P(z^{(i)}) && (8)\\
&= \sum_{z^{(i)}=1}^{3} P(w^{(i)}|z^{(i)})\phi_1^{\mathbb{1}\{z^{(i)}=1\}}\phi_2^{\mathbb{1}\{z^{(i)}=2\}}\phi_1^{\{z^{(i)}=3\}} && (9)\\
&= \phi_1\delta(w^{(i)}) + \phi_2 Exp(w^{(i)}) + \phi_3\mathcal{N}(w^{(i)}; \mu, \sigma^2), && (10)
\end{aligned}
$$

where, by introducing the δ I have used the so-called "generalized PDF" of the indicator function PMF.

1.1 The Expectation and Variance of W_{tot}^{DMA}

Letting:

$$g_1(\mu, \sigma) \equiv \int_0^\infty x\mathcal{N}(x; \mu, \sigma^2)dx = \frac{\mu}{2}\left[\text{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + 1\right] + \frac{\sigma\exp(-\mu^2/(2\sigma^2))}{\sqrt{2\pi}}, \quad (11)$$

and

$$g_2(\mu, \sigma) \equiv \int_0^\infty x^2\mathcal{N}(x; \mu, \sigma^2)dx = \frac{\mu^2 + \sigma^2}{2}\left[\text{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + 1\right] + \frac{\mu\sigma\exp(-\mu^2/(2\sigma^2))}{\sqrt{2\pi}}, \quad (12)$$

$E[W^{(i)}]$ and $E[W^{(i)2}]$ can be easily calculated:

$$E[W^{(i)}] = \phi_2 \int_0^\infty x\lambda e^{-\lambda x} dx + \phi_3 \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} x \exp\left[\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx \tag{13}$$

$$= \left(\frac{\phi_2}{\lambda} + \phi_3 g_1(\mu,\sigma)\right), \tag{14}$$

and

$$E[W^{(i)2}] = \phi_2 \int_0^\infty x^2\lambda e^{-\lambda x} dx + \phi_3 \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} x^2 \exp\left[\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] dx \tag{15}$$

$$= \phi_2 \frac{2}{\lambda^2} + \phi_3 g_2(\mu,\sigma). \tag{16}$$

The variance in weighted minutes viewed for the i^{th} person is thus:

$$Var[W^{(i)}] = E[W^{(i)2}] - E[W^{(i)}]^2 \tag{17}$$

$$= \phi_2 \frac{2}{\lambda^2} + \phi_3 g_2(\mu,\sigma) - \left(\frac{\phi_2}{\lambda} + \phi_3 g_1(\mu,\sigma)\right)^2. \tag{18}$$

To compute the expectation for the total weighted minutes viewed, I use the above expressions and the law of iterated expectations (conditioning on N_{DMA}):

$$E[W_{tot}^{DMA}] = E\left[\sum_{i=1}^{N_{DMA}} W^{(i)}\right] \tag{19}$$

$$= E\left[E\left[\sum_{i=1}^{N_{DMA}} W^{(i)} \middle| N_{DMA}\right]\right] \tag{20}$$

$$= E\left[N_{DMA} E[W^{(i)}]\right] \tag{21}$$

$$= \langle N_{DMA}\rangle E[W^{(i)}] \tag{22}$$

$$= \langle N_{DMA}\rangle \left(\frac{\phi_2}{\lambda} + \phi_3 g_1(\mu,\sigma)\right). \tag{23}$$

Similarly, the variance can be computed using the law of total variance:

$$Var[W_{tot}^{DMA}] = Var\left[\sum_{i=1}^{N_{DMA}} W^{(i)}\right] \tag{24}$$

$$= E\left[Var\left[\sum_{i=1}^{N_{DMA}} W^{(i)}\right] \middle| N_{DMA}\right] + Var\left[E\left[\sum_{i=1}^{N_{DMA}} W^{(i)}\right] \middle| N_{DMA}\right] \tag{25}$$

$$= E\left[N_{DMA} Var\left[W^{(i)}\right]\right] + Var\left[N_{DMA} E[W^{(i)}]\right] \tag{26}$$

$$= E[N_{DMA}]Var[W^{(i)}] + Var[N_{DMA}]E[W^{(i)}]^2 \tag{27}$$

$$= \langle N_{DMA}\rangle \left[\phi_2 \frac{2}{\lambda^2} + \phi_3 g_2(\mu,\sigma) - \left(\frac{\phi_2}{\lambda} + \phi_3 g_1(\mu,\sigma)\right)^2\right] + \sigma^2_{N_{DMA}} \left[\phi_2 \frac{2}{\lambda^2} + \phi_3 g_2(\mu,\sigma)\right]^2, \tag{28}$$

where I have used the fact that all $W^{(i)}s$ are independent.

2 A Generative Model with Promos

I can construct a generative model that takes into account the presence of promos by slightly modifying the generative model of the previous section. The idea is that before $Z^{(i)}$ is drawn which determines the class of the i^{th} respondent, the respondent is exposed to promos a certain number of times, $N^{(i)}$, (drawn from a Poisson distribution). The more promos the person is exposed to, the higher probability that the person converts from class 1 to either class 2 or 3. I can take this into account by defining a function $\Delta(n)$ which increases monotonically from 0 to 1 as n goes from 0 to ∞. I then modify the probability that a person belongs to class 1 by subtracting this amount from the probability that the person is class 1, and I modify the probabilities of the other 2 classes by adding half this amount.

The generative model is thus:

$$N_{DMA} \sim \mathcal{D}(\Theta) \tag{29}$$

$$N^{(i)} \sim Poiss(\alpha) \tag{30}$$

$$Z_{pr}^{(i)}|N^{(i)} = n^{(i)} \sim Cat\left(\phi_1(1-\Delta(n^{(i)})), \phi_2 + \Delta(n^{(i)})\phi_1/2, \phi_3 + \Delta(n^{(i)})\phi_1/2\right) \tag{31}$$

$$\begin{cases} W^{(i)}|Z_{pr}^{(i)} = 1 \sim \mathbb{1}\{w=0\} \\ W^{(i)}|Z_{pr}^{(i)} = 2 \sim Exp(\lambda) \\ W^{(i)}|Z_{pr}^{(i)} = 3 \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \tag{32}$$

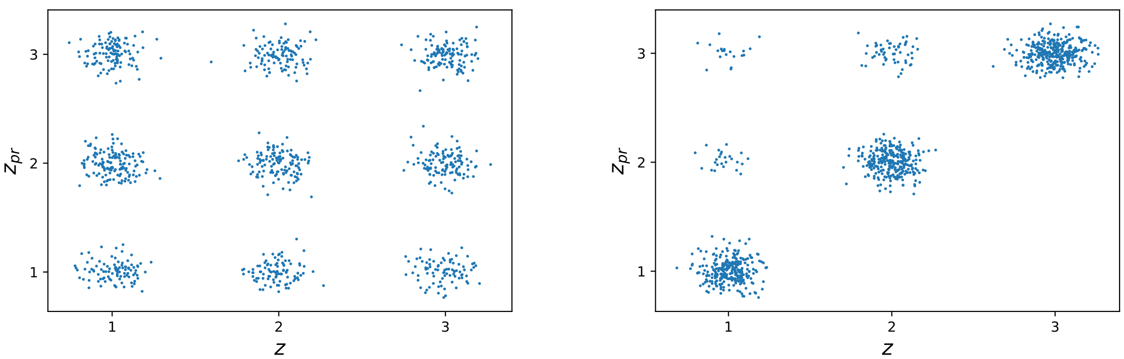
$$W_{tot}^{DMA} = \sum_{i=1}^{N_{DMA}} W^{(i)}, \tag{33}$$

for $i = 1, 2, \ldots, N_{DMA}$. As a reminder, the PMF of a Poisson distribution is:

$$P(N = n; \alpha) = \frac{\alpha^n e^{-\alpha}}{n!} \tag{34}$$

The problem with this model is that there is no correlation between $Z^{(i)}$ and $Z_{pr}^{(i)}$ since they are independent. Intuitively, one would expect, that, at least for a moderate amount of promo exposures, for the i^{th} person, if $Z^{(i)} = z^{(i)}$ it would be likely that $Z_{pr}^{(i)} = z^{(i)}$ since, after all, this is the same person, behaving similarly.

I incorporate this correlation into a slightly more complex generative model, where I first draw $Z^{(i)}$, then condition $Z_{pr}^{(i)}$ on that value (as well as $N^{(i)}$). In words, the model is such that if $N^{(i)} = 0$, then $Z_{pr}^{(i)} = Z^{(i)}$ with 100% probability (i.e., no change in behavior since the person saw no promos). For $N^{(i)} > 0$, if $Z^{(i)} = 1$, then $Z_{pr}^{(i)} \sim Cat(1-\Delta(n^{(i)}), \Delta(n^{(i)})/2, \Delta(n^{(i)})/2)$ (that is, you have a small chance of getting to the second or third class from the first). If $Z^{(i)} = 2$, $Z_{pr}^{(i)} \sim Cat(0, 1-\Delta(n^{(i)}), \Delta(n^{(i)}))$ (you have a small chance of getting to the third class from the second). Finally, if $Z^{(i)} = 3$, you are already in the highest viewing category, so $Z_{pr}^{(i)} = 3$ with 100% probability. Note that in this model, exposures to promos will only help push you to a higher viewing category, and not a lower one.



(a) Generative Model 1

(b) Generative Model 2

Figure 2: A simulation for both generative models with $N_{DMA} = 1000$, using a $Poiss(4)$ distribution, and using $\Delta(n^{(i)}) = 1 - e^{-n^{(i)}/30}$. Guassian jitter has been added to each data point for the purposes of visualization.

The generative model is:

$$
\begin{aligned}
&N_{DMA} \sim \mathcal{D}(\Theta) \\
&N^{(i)} \sim Poiss(\alpha) \\
&Z^{(i)} \sim Cat(\phi_1, \phi_2, \phi_3) \\
&\begin{cases}
Z_{pr}^{(i)} | N^{(i)} = 0, Z^{(i)} = z^{(i)} \sim Cat(\boldsymbol{I}_{z^{(i)},*}) \\
Z_{pr}^{(i)} | N^{(i)} = n^{(i)} > 0, Z^{(i)} = 1 \sim Cat(1 - \Delta(n^{(i)}), \Delta(n^{(i)})/2, \Delta(n^{(i)})/2) \\
Z_{pr}^{(i)} | N^{(i)} = n^{(i)} > 0, Z^{(i)} = 2 \sim Cat(0, 1 - \Delta(n^{(i)}), \Delta(n^{(i)})) \\
Z_{pr}^{(i)} | N^{(i)} = n^{(i)} > 0, Z^{(i)} = 3 \sim Cat(0, 0, 1)
\end{cases} \\
&\begin{cases}
W^{(i)} | Z_{pr}^{(i)} = 1 \sim \mathbb{1}\{w = 0\} \\
W^{(i)} | Z_{pr}^{(i)} = 2 \sim Exp(\lambda) \\
W^{(i)} | Z_{pr}^{(i)} = 3 \sim \mathcal{N}(\mu, \sigma^2)
\end{cases} \\
&W_{tot}^{DMA} = \sum_{i=1}^{N_{DMA}} W^{(i)},
\end{aligned}
$$

for $i = 1, 2, \ldots, N_{DMA}$. Here $\boldsymbol{I}_{z^{(i)},*}$ is the $z^{(i)th}$ row of the 3×3 identity matrix.

The notation can be made slightly tighter by introducing a transition matrix, $\boldsymbol{T}_n$, defined such that $P(Z_{pr} = z_{pr} | N = n, Z = z) = (\boldsymbol{T}_n)_{z, z_{pr}}$:

$$
\boldsymbol{T}_n \equiv \begin{bmatrix} 1 - \Delta(n) & \Delta(n)/2 & \Delta(n)/2 \\ 0 & 1 - \Delta(n) & \Delta(n) \\ 0 & 0 & 1 \end{bmatrix}.
$$

Notice that, as it should be, $\boldsymbol{T}_n$ become the identity matrix when $n = 0$. With this definition, the generative model is:

$$
\begin{aligned}
&N_{DMA} \sim \mathcal{D}(\Theta) \\
&N^{(i)} \sim Poiss(\alpha) \\
&Z^{(i)} \sim Cat(\phi_1, \phi_2, \phi_3) \\
&Z_{pr}^{(i)} | N^{(i)} = n^{(i)}, Z^{(i)} = z^{(i)} \sim Cat\left((\boldsymbol{T}_{n^{(i)}})_{z^{(i)},*}\right) \\
&\begin{cases} W^{(i)} | Z_{pr}^{(i)} = 1 \sim \mathbb{1}\{w = 0\} \\ W^{(i)} | Z_{pr}^{(i)} = 2 \sim Exp(\lambda) \\ W^{(i)} | Z_{pr}^{(i)} = 3 \sim \mathcal{N}(\mu, \sigma^2) \end{cases} \\
&W_{tot}^{DMA} = \sum_{i=1}^{N_{DMA}} W^{(i)}.
\end{aligned}
$$

Conditioning $Z_{pr}^{(i)}$ on $Z^{(i)}$ fixes this correlation issue by making $Z_{pr}^{(i)}$ dependent on $Z^{(i)}$. In a strict sense, dependence between random variables does not imply correlation, and further, what is even meant by correlation between nominal categorical variables (as in this case) is unclear. Nevertheless, to illustrate how this model incorporates "correlation" between $Z_{pr}^{(i)}$ and $Z^{(i)}$, I run a simulation of the first generative model with $N_{DMA} = 1000$, $\alpha = 4$ (i.e., on average people were exposed to 4 promos) and with $\Delta(n^{(i)}) = 1 - e^{-n^{(i)}/30}$. A value of 30 in the denominator of the exponential means that at $n = 5$, this corresponds to $\Delta \approx 15\%$; i.e., after 5 exposures, you are 15% more likely to convert from class 1 to either class 2 or 3 (which seems reasonable).

I show the results of this simulation in Fig. 2a. I make a scatter plot of all $(z^{(i)}, z_{pr}^{(i)})$ pairs drawn. For visualization purposes, for each point on the plot, I add a bit of Gaussian noise. In this model, although the probability of Z_{pr} being class 1 is lower (which is what we wanted the promo model to capture), there is very little "correlation" amongst the data points. I perform the same simulation for the second generative model and show the results in in Fig. 2b. There is clear correlation between the variables, shown by the high concentration of points at $(1,1)$, $(2,2)$ and $(3,3)$. There are no points at $(2,1)$, $(3,1)$ and $(3,2)$ since I have built into the model that it is impossible to switch to a lower viewing class after being exposed to a promo. From a business perspective, the small concentration of points at $(1,2)$, $(1,3)$ and $(2,3)$ are the valuable customers we have converted by showing promos. In particular the cluster of points at $(1,3)$ is very valuable, as these are customers who have converted from the the zero-minute class (the lowest viewing class) to the Gaussian class (the highest viewing class).

Given this generative model, I now compute $P(W^{(i)} = w^{(i)})$ so that I may compute the desired expectations and variances later on. To lighten up notation, I drop all (i) superscripts:

$$
\begin{aligned}
P(w) &= \sum_z \sum_{z_{pr}} \sum_n P(w, n, z_{pr}, z) \\
&= \sum_z P(z) \sum_{z_{pr}} P(w|z_{pr}) \sum_n P(z_{pr}|n, z) P(n) \\
&= \sum_z \phi_1^{\mathbb{1}\{z=1\}} \phi_2^{\mathbb{1}\{z=2\}} \phi_3^{\mathbb{1}\{z=3\}} \\
&\times \sum_{z_{pr}} \left[\mathbb{1}\{z_{pr}=1\}\delta(w) + \mathbb{1}\{z_{pr}=2\} Exp(w;\lambda) + \mathbb{1}\{z_{pr}=3\}\mathcal{N}(w;\mu,\sigma^2)\right] \\
&\times \left\{ \mathbb{1}\{z_{pr}=z\}e^{-\alpha} + \sum_{n=1}^{\infty} P(n) \left[\mathbb{1}\{z=1\}[1-\Delta(n)]^{\mathbb{1}\{z_{pr}=1\}} \left(\frac{\Delta(n)}{2}\right)^{\mathbb{1}\{z_{pr}=2\}} \left(\frac{\Delta(n)}{2}\right)^{\mathbb{1}\{z_{pr}=3\}} \right.\right. \\
&\left.\left. + \mathbb{1}\{z=2\}\mathbb{1}\{z_{pr}\neq 1\}[1-\Delta(n)]^{\mathbb{1}\{z_{pr}=2\}} \Delta(n)^{\mathbb{1}\{z_{pr}=3\}} + \mathbb{1}\{z=3\}\mathbb{1}\{z_{pr}=3\} \right] \right\} \\
&= \sum_z \phi_1^{\mathbb{1}\{z=1\}} \phi_2^{\mathbb{1}\{z=2\}} \phi_3^{\mathbb{1}\{z=3\}} \\
&\times \sum_{z_{pr}} \left[\mathbb{1}\{z_{pr}=1\}\delta(w) + \mathbb{1}\{z_{pr}=2\} Exp(w;\lambda) + \mathbb{1}\{z_{pr}=3\}\mathcal{N}(w;\mu,\sigma^2)\right] \\
&\times \left\{ \mathbb{1}\{z_{pr}=z\}e^{-\alpha} + \mathbb{1}\{z=1\}[(1-e^{-\alpha}) - \bar{\Delta}]^{\mathbb{1}\{z_{pr}=1\}} \left(\frac{\bar{\Delta}}{2}\right)^{\mathbb{1}\{z_{pr}=2\}} \left(\frac{\bar{\Delta}}{2}\right)^{\mathbb{1}\{z_{pr}=3\}} \right. \\
&\left. + \mathbb{1}\{z=2\}\mathbb{1}\{z_{pr}\neq 1\}[(1-e^{-\alpha}) - \bar{\Delta}]^{\mathbb{1}\{z_{pr}=2\}} \bar{\Delta}^{\mathbb{1}\{z_{pr}=3\}} + \mathbb{1}\{z=3\}\mathbb{1}\{z_{pr}=3\}(1-e^{-\alpha}) \right\} \\
&= \sum_z \phi_1^{\mathbb{1}\{z=1\}} \phi_2^{\mathbb{1}\{z=2\}} \phi_3^{\mathbb{1}\{z=3\}} \\
&\times \left\{ \delta(w)\left[\mathbb{1}\{z=1\}e^{-\alpha} + \mathbb{1}\{z=1\}((1-e^{-\alpha}) - \bar{\Delta})\right] \right. \\
&+ Exp(w;\lambda) \left[\mathbb{1}\{z=2\}e^{-\alpha} + \mathbb{1}\{z=1\}\left(\frac{\bar{\Delta}}{2}\right) + \mathbb{1}\{z=2\}((1-e^{-\alpha}) - \bar{\Delta}) \right] \\
&\left. + \mathcal{N}(w;\mu,\sigma^2) \left[\mathbb{1}\{z=3\}e^{-\alpha} + \mathbb{1}\{z=1\}\left(\frac{\bar{\Delta}}{2}\right) + \mathbb{1}\{z=2\}\bar{\Delta} + \mathbb{1}\{z=3\}(1-e^{-\alpha}) \right] \right\},
\end{aligned}
$$

so that finally, we have:

$$
\begin{aligned}
P(w^{(i)}) &= \phi_1\left(1-\bar{\Delta}\right)\delta(w) + \phi_2\left(1-\bar{\Delta}\right) Exp(w^{(i)};\lambda) + \phi_3 \mathcal{N}(w^{(i)};\mu,\sigma) \\
&+ \phi_1 \left(\frac{\bar{\Delta}}{2}\right)\left[Exp(w^{(i)};\lambda) + \mathcal{N}(w^{(i)};\mu,\sigma)\right] + \phi_2 \bar{\Delta} \mathcal{N}(w^{(i)};\mu,\sigma). \qquad (35)
\end{aligned}
$$

Here, I have first marginalized over $Z^{(i)}$, $Z_{pr}^{(i)}$ and $n^{(i)}$, used the chain rule of probability, cancelled out some variables due to conditional independence and then plugged in the actual formulas for the various PMFs. I have also heavily relied on indicator functions, which makes the summations significantly easier to evaluate. I have also defined: $\sum_{n^{(i)}=0}^{\infty} P(n^{(i)})\Delta(n^{(i)}) \equiv \bar{\Delta}$, the average (across number of promos viewed) probability that a person will transition from a lower viewing class to a higher viewing class. This can roughly be thought of as the conversion rate averaged across number of promos seen. I have also used the "generalized PDF" to incorporate the $\mathbb{1}\{w=0\}$ PMF into this PDF. Also note that, the math works out such that the results only depend on $\bar{\Delta}$, and do not explicitly depend on α, or even the probability distribution used for $N^{(i)}$ for that matter.

As a sanity check, it is not difficult to verify that this distribution integrates to unity.

2.1 The Expectation and Variance of W_{tot}^{DMA}

Now that $P(w^{(i)})$ is known, computing the $E[W^{(i)}]$ and $E[W^{(i)2}]$ under this model is easy:

$$E[W^{(i)}] = \frac{\phi_2(1-\bar{\Delta})}{\lambda} + \phi_3 g_1(\mu,\sigma) + \bar{\Delta}\left[\frac{\phi_1}{2}\left(\frac{1}{\lambda} + g_1(\mu,\sigma)\right) + \phi_2 g_1(\mu,\sigma)\right] \tag{36}$$

and

$$E[W^{(i)2}] = \phi_2(1-\bar{\Delta})\frac{2}{\lambda^2} + \phi_3 g_2(\mu,\sigma) + \bar{\Delta}\left[\frac{\phi_1}{2}\left(\frac{2}{\lambda^2} + g_2(\mu,\sigma)\right) + \phi_2 g_2(\mu,\sigma)\right]. \tag{37}$$

As usual:

$$Var[W^{(i)}] = E[W^{(i)2}] - E[W^{(i)}]^2, \tag{38}$$

where the formulas for the terms on the right are defined above.

Computing the expectation and variance of W_{tot}^{DMA} is very similar to computing the corresponding value from the previous section, and the formulas are given by:

$$E[W_{tot}^{DMA}] = \langle N_{DMA}\rangle E[W^{(i)}] \tag{39}$$

and

$$Var[W_{tot}^{DMA}] = \langle N_{DMA}\rangle Var[W^{(i)}] + \sigma_{N_{DMA}}^2 E[W^{(i)}]^2, \tag{40}$$

where all formulas necessary to compute these values are given above. It is convenient that through the math, all free parameters have collapsed into just a single free parameter, $\bar{\Delta}$.

3 Covariance Computation and Expected lift Due to Promos

In this section, I use the model to compute $E[W_{tot,\bar{\Delta}}^{DMA}/W_{tot,0}^{DMA}]$, that is, the expected lift in total weighted minutes viewed under this model relative to when no promos are shown (the 0 subscript indicates $\bar{\Delta}=0$ and the $\bar{\Delta}$ subscript indicates $\bar{\Delta}>0$).

To compute this value, I will first need to compute the covariance between $W_{tot,\bar{\Delta}}^{DMA}$ and $W_{tot,0}^{DMA}$. Using the law of total covariance, I have that:

$$
\begin{aligned}
Cov\left[W_{tot,0}^{DMA}, W_{tot,\bar{\Delta}}^{DMA}\right] &= Cov\left[\sum_{i=1}^{N_{DMA}} W_0^{(i)}, \sum_{j=1}^{N_{DMA}} W_{\bar{\Delta}}^{(j)}\right] \\
&= E\left[Cov\left[\sum_{i=1}^{N_{DMA}} W_0^{(i)}, \sum_{j=1}^{N_{DMA}} W_{\bar{\Delta}}^{(j)} \middle| N_{DMA}\right]\right] \\
&+ Cov\left[E\left[\sum_{i=1}^{N_{DMA}} W_0^{(i)} \middle| N_{DMA}\right], E\left[\sum_{j=1}^{N_{DMA}} W_{\bar{\Delta}}^{(j)} \middle| N_{DMA}\right]\right] \\
&= E\left[\sum_{i=1}^{N_{DMA}} \sum_{j=1}^{N_{DMA}} Cov\left[W_0^{(i)}, W_{\bar{\Delta}}^{(j)}\right]\right] + Cov\left[N_{DMA}E\left[W_0^{(i)}\right], N_{DMA}E\left[W_{\bar{\Delta}}^{(j)}\right]\right] \\
&= E\left[N_{DMA}Cov\left[W_0^{(i)}, W_{\bar{\Delta}}^{(i)}\right]\right] + E[W_0^{(i)}]E[W_{\bar{\Delta}}^{(j)}]Cov[N_{DMA}, N_{DMA}] \\
&= \langle N_{DMA}\rangle Cov\left[W_0^{(i)}, W_{\bar{\Delta}}^{(i)}\right] + \sigma_{N_{DMA}}^2 E[W_0^{(i)}]E[W_{\bar{\Delta}}^{(j)}],
\end{aligned}
$$

where I have used the fact that since $W_0^{(i)}$ and $W_{\bar{\Delta}}^{(j)}$ are independent, their covariance is zero.

What is left to do is to compute $Cov\left[W_0^{(i)}, W_{\bar{\Delta}}^{(i)}\right]$, and to do this I first compute $P(z^{(i)}, z_{pr}^{(i)})$:

$$
\begin{aligned}
P(z^{(i)}, z_{pr}^{(i)}) &= \sum_{n^{(i)}} P(z_{pr}^{(i)}|z^{(i)}, n^{(i)})P(z^{(i)})P(n^{(i)}) \\
&= \sum_{n^{(i)}} P(n^{(i)})(\boldsymbol{T}_{n^{(i)}})_{z^{(i)}, z_{pr}^{(i)}} P(z^{(i)}) \\
&= (\bar{\boldsymbol{T}}_{n^{(i)}})_{z^{(i)}, z_{pr}^{(i)}} \phi_1^{\mathbb{1}\{z^{(i)}=1\}} \phi_1^{\mathbb{1}\{z^{(i)}=2\}} \phi_3^{\mathbb{1}\{z^{(i)}=3\}},
\end{aligned}
$$

where the bar over the matrix element denotes an average over n for that element. For example $(\bar{\boldsymbol{T}}_n)_{11} = 1 - \bar{\Delta}$.

Defining the random variables, $W_1 \sim \mathbb{1}\{w = 0\}$, $W_2 \sim Exp(\lambda)$, and $W_3 \sim \mathcal{N}(\mu, \sigma^2)$, I may now use the law of total expectation to compute $E[W_0^{(i)} W_{\bar{\Delta}}^{(i)}]$:

$$
\begin{aligned}
E[W_0^{(i)} W_{\bar{\Delta}}^{(i)}] &= \sum_{z^{(i)}, z_{pr}^{(i)}} E[W_0^{(i)} W_{\bar{\Delta}}^{(i)} | Z^{(i)} = z^{(i)}, Z_{pr}^{(i)} = z_{pr}^{(i)}] P(z_{pr}^{(i)}, z^{(i)}) \\
&= \sum_{z^{(i)}, z_{pr}^{(i)}} E[W_{z^{(i)}} W_{z_{pr}^{(i)}} | Z^{(i)} = z^{(i)}, Z_{pr}^{(i)} = z_{pr}^{(i)}] (\bar{\boldsymbol{T}}_{n^{(i)}})_{z^{(i)}, z_{pr}^{(i)}} \phi_1^{\mathbb{1}\{z^{(i)}=1\}} \phi_1^{\mathbb{1}\{z^{(i)}=2\}} \phi_3^{\mathbb{1}\{z^{(i)}=3\}} \\
&= \cancel{E[W_1^{(i)} W_1^{(i)}]\phi_1(1-\bar{\Delta})} + \cancel{E[W_1^{(i)} W_2^{(i)}]\phi_1\left(\frac{\bar{\Delta}}{2}\right)} + \cancel{E[W_1^{(i)} W_3^{(i)}]\phi_1\left(\frac{\bar{\Delta}}{2}\right)} \\
&+ E[W_2^{(i)} W_2^{(i)}]\phi_2(1-\bar{\Delta}) + E[W_2^{(i)} W_3^{(i)}]\phi_2\bar{\Delta} + E[W_3^{(i)} W_3^{(i)}]\phi_3 \\
&= \phi_2 \frac{2}{\lambda^2}(1-\bar{\Delta}) + \phi_2 \frac{g_1(\mu, \sigma)\bar{\Delta}}{\lambda} + \phi_3 g_2(\mu, \sigma).
\end{aligned}
$$

Finally, the covariance is given by the standard formula:

$$
Cov\left[W_0^{(i)}, W_{\bar{\Delta}}^{(i)}\right] = E[W_0^{(i)} W_{\bar{\Delta}}^{(i)}] - E[W_0^{(i)}]E[W_{\bar{\Delta}}^{(i)}] \tag{41}
$$

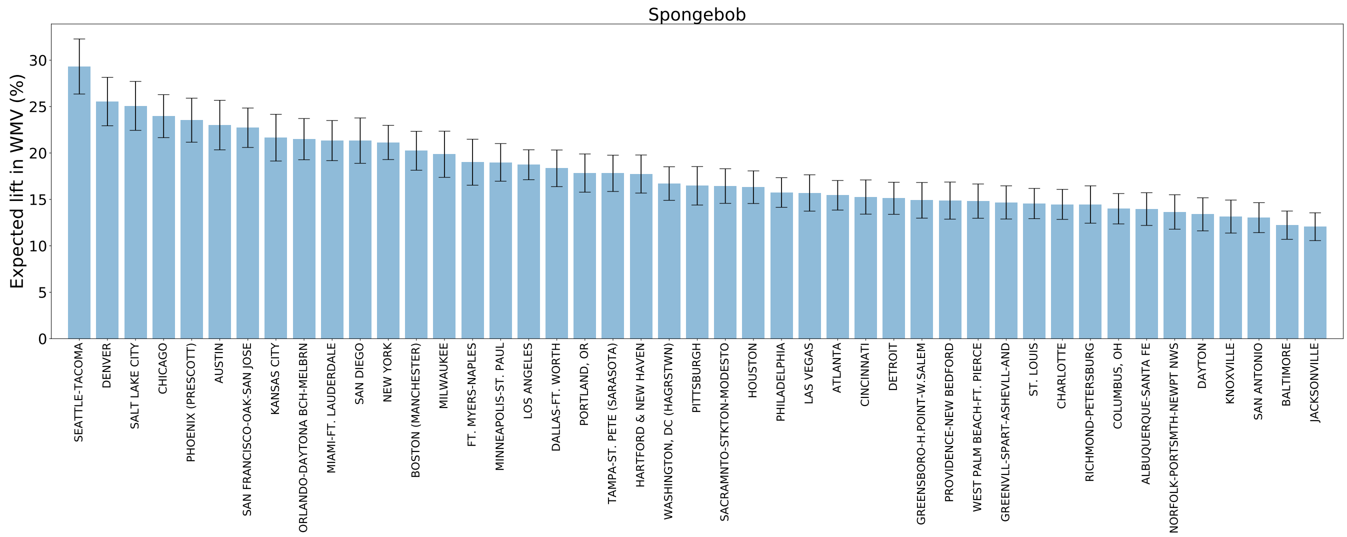


Figure 3: Expected lift for Spongebob for each DMA ranked by lift.

Finally, in order to compare the percent lift across DMAs, I compute the expected value of the lift (as well the variance of this to get a sense of the uncertainty). In order to compute these values, I use a second-order accurate Taylor expansion[1]:

$$E\left[\frac{W^{DMA}_{tot,\bar{\Delta}}}{W^{DMA}_{tot,0}} - 1\right] \approx \frac{E[W^{DMA}_{tot,\bar{\Delta}}]}{E[W^{DMA}_{tot,0}]} - \frac{Cov[W^{DMA}_{tot,0}, W^{DMA}_{tot,\bar{\Delta}}]}{E[W^{DMA}_{tot,0}]^2} + \frac{Var[W^{DMA}_{tot,0}]E[W^{DMA}_{tot,\bar{\Delta}}]}{E[W^{DMA}_{tot,0}]^3} - 1 \tag{42}$$

$$Var\left[\frac{W^{DMA}_{tot,\bar{\Delta}}}{W^{DMA}_{tot,0}} - 1\right] \approx \frac{E[W^{DMA}_{tot,\bar{\Delta}}]^2}{E[W^{DMA}_{tot,0}]^2}\left(\frac{Var[W^{DMA}_{tot,\bar{\Delta}}]}{E[W^{DMA}_{tot,\bar{\Delta}}]^2} - 2\frac{Cov[W^{DMA}_{tot,0}, W^{DMA}_{tot,\bar{\Delta}}]}{E[W^{DMA}_{tot,0}]E[W^{DMA}_{tot,\bar{\Delta}}]} + \frac{Var[W^{DMA}_{tot,0}]}{E[W^{DMA}_{tot,0}]^2}\right) \tag{43}$$

All expectations, variances, and covariances in these equations have been explicitly computed within these notes. I use the ML estimates of ϕ_1, ϕ_2, ϕ_3, μ, σ and λ fitted separately for each DMA. To compute $E[N_{DMA}]$ and σ^2_{DMA} I fit time series to the number of respondents in each DMA using Facebook's prophet model and I forecast to the desired day (the day when we want the promo to be shown).

I show an example of the expected lift in weighted minutes viewed for Spongebob in Fig. 3 using a value of $\bar{\Delta} = 0.005$ (an average conversion rate of 0.5%).

[1]For example, see `http://www.stat.cmu.edu/~hseltman/files/ratio.pdf`