2020 March 9

Scribed by Roh

hyunwoo@uchicago.edu

1 Answer to student question in Office Hour

When X is binary variable, how can we express β? In general setting,

$$\beta = \frac{Cov(X,Y)}{\text{Var}(X)}$$

where

$$Cov(X,Y) = E[XY] - E[X]E[Y] \tag{1}$$

Note that

$$\begin{aligned} E[XY] &= E[XE[Y|X]] \\ &= 1 \cdot E[Y|X=1] \cdot P\{X=1\} + 0 \cdot E[Y|X=0] \cdot P\{X=0\} \\ &= P(X=1)\, E[Y|X=1] \end{aligned}$$

Similarly,

$$\begin{aligned} E[X]E[Y] &= [1 \cdot P(X=1) + 0 \cdot P(X=0)] \times E[E[Y|X]] \\ &= P(X=1) \times E[E[Y|X]] \\ &= P(X=1) \times [E[Y|X=1] \cdot P(X=1) + E[Y|X=0] \cdot P(X=0)] \\ &= P(X=1)^2 \cdot E[Y|X=1] + P(X=1) \cdot P(X=0) \cdot E[Y|X=0] \end{aligned}$$

Now we put these two back into equation (1) above to get

$$\begin{aligned} Cov(X,Y) &= E[XY] - E[X]E[Y] \\ &= P(X=1)\, E[Y|X=1] - P(X=1)^2 \cdot E[Y|X=1] + P(X=1) \cdot P(X=0) \cdot E[Y|X=0] \\ &= \left(P(X=1) - P(X=1)^2\right) \cdot E[Y|X=1] - P(X=1)\,(1 - P(X=1)) \cdot E[Y|X=0] \\ &= \left(P(X=1) - P(X=1)^2\right) \cdot (E[Y|X=1] - E[Y|X=0]) \end{aligned}$$

Since

$$\begin{aligned} \text{Var}[X] &= E[X^2] - E[X]^2 \\ &= P(X=1) - P(X=1)^2 \end{aligned}$$

we have

$$\begin{aligned} \beta &= \frac{Cov(X,Y)}{\text{Var}(X)} \\ &= \frac{\left(P(X=1) - P(X=1)^2\right)(E[Y|X=1] - E[Y|X=0])}{P(X=1) - P(X=1)^2} \\ &= E[Y|X=1] - E[Y|X=0] \end{aligned}$$

2 Continue from Mar 6

Especially how do we calculate the following?

$$\frac{\delta}{\delta p(y_p)} \int_{-\infty}^{\infty} p(y) \log \frac{p(y)}{p_0(y)} dy$$

We use the functional derivative to get (By Euler-Lagrange Equation)

$$\begin{aligned}
&\frac{\delta}{\delta p(y)} \int \left(p(y) \log \frac{p(y)}{p_0(y)} \right) dy \\
=&\frac{\delta}{\delta p(y)} \int \left(p(y) \log p(y) - p(y) \log p_0(y) \right) dy \\
=&\frac{\delta}{\delta p(y)} \int \left(p(y) \log p(y) \right) dy - \frac{\delta}{\delta p(y)} \int \left(p(y) \log p_0(y) \right) dy
\end{aligned}$$

To use the Euler-Lagrange equation, note that we can write

$$F[p(y)] = \int L\left(y, p(y)\right) dy$$

To continue with the first term where $L(y, p(y) = p(y) \log p(y)$,

$$\begin{aligned}
\frac{\delta F}{\delta f(x)} &= \frac{\partial L}{\partial f(x)} - \frac{d}{dx} \frac{\partial L}{\partial f'(x)} \\
\Rightarrow \frac{\delta F[p(y)]}{\delta p(y)} &= \frac{\partial L}{\partial f(x)} - 0 \qquad \because \text{we don't have derivative of } p'(y) \text{ here} \\
&= p(y) \frac{1}{p(y)} + \log p(y) \\
&= 1 + \log p(y)
\end{aligned}$$

For the second term where $L(y, p(y) = p(y) \log p_0(y)$,

$$\begin{aligned}
\frac{\delta F}{\delta f(x)} &= \frac{\partial L}{\partial f(x)} - \frac{d}{dx} \frac{\partial L}{\partial f'(x)} \\
\Rightarrow \frac{\delta F[p(y)]}{\delta p(y)} &= \frac{\partial L}{\partial f(x)} - 0 \qquad \because \text{we don't have derivative of } p'(y) \text{ here} \\
&= \log p_0(y)
\end{aligned}$$

Plug these two back to get

$$\begin{aligned}
&\frac{\delta}{\delta p(y_p)} \int \left(p(y) \log \frac{p(y)}{p_0(y)} \right) dy \\
=&1 + \log p(y) - \log p_0(y) \\
=&1 + \log \frac{p(y)}{p_0(y)}
\end{aligned}$$

We could do this using the definition of functional derivative instead of using Euler-Lagrange equation as follows:

$$\frac{\delta}{\delta p(y_p)} \underbrace{\int (p(y)\log p(y))\,dy}_{\equiv G[p]}$$

$$\begin{aligned}
\text{where} G[p] &= \int g(y,p(y))dy \\
&= \frac{\delta G[p]}{\delta p(y_p)} \\
&= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int \left[g\left(y,p(y)+\epsilon\delta(y-y_p)\right)-g(y,p(y))\right]dy
\end{aligned}$$

Since we are only interested at terms in the integrand that are first order in ϵ,
we can Taylor-exapand$g\left(y,p(y)+\epsilon\delta(y-y_p)\right)$ to first order to get

$$\begin{aligned}
\frac{\delta G[p]}{\delta p(y_p)} &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int\left[g\left(y,p(y)+\epsilon\frac{\partial g\left(y,p(y)\right)}{\partial p(y)}\delta(y-y_p)\right)-g(y,p(y))\right]dy \\
&= \int \frac{\partial g\left(y,p(y)\right)}{\partial p(y)}\delta(y-y_p)dy \\
&= \frac{\partial g\left(y_p,p(y)\right)}{\partial p(y)} = 1+\log p(y)
\end{aligned}$$

The equality of last line holds because we are assuming that in integrals of the form $\int p(y)\delta(y-y_p)dy$ that y_p is always within the limits of integration so that the result is $\int p(y)\delta(y-y_p)dy = p(y_p)$. In all cases, if x is outside the limits of integration, then the integral evaluates to zero.

2.1 Back to the main problem

Recall from the Mar_6 note, we did functional derivative with respect to $p(y)$ to get

$$\frac{\partial\mathcal{L}}{\partial p} = \log\frac{p(y)}{p_0(y)} + 1 - \lambda_1 F_1(y) - \cdots - \lambda_m F_m(y) = 0$$

After resolving to $p(y)$ and normalizing the result, we arrive at the **Boltzmann distribution** which is

$$p_B(y) = \frac{1}{Z}p_0(y)\exp\left\{\lambda_1 F_1(y) + \cdots + \lambda_m F_m(y)\right\}$$

with the normalizing constant is there to make sure the integral over distribution equals to 1 as follows:

$$Z(\lambda) = \int_{-\infty}^{\infty} p_0(y)\exp\left\{\lambda_1 F_1(y) + \cdots + \lambda_m F_m(y)\right\}dy$$

so that the following satisfies

$$\int_{-\infty}^{\infty} p_B(y)dy = 1$$

The partial derivatives of L with respective to λ read As we already have calculated our normalized solution to $p(y)$, which $p_B(y)$, we can insert this result into the derivatives:

$$\frac{\partial L}{\partial\lambda_k} = \int_{-\infty}^{\infty} F_k(y)\underbrace{\frac{1}{Z}p_0(y)\exp\left\{\lambda_1 F_1(y) + \cdots + \lambda_m F_m(y)\right\}}_{P_B(y)}dy - f_k = 0$$

for m different constraints.

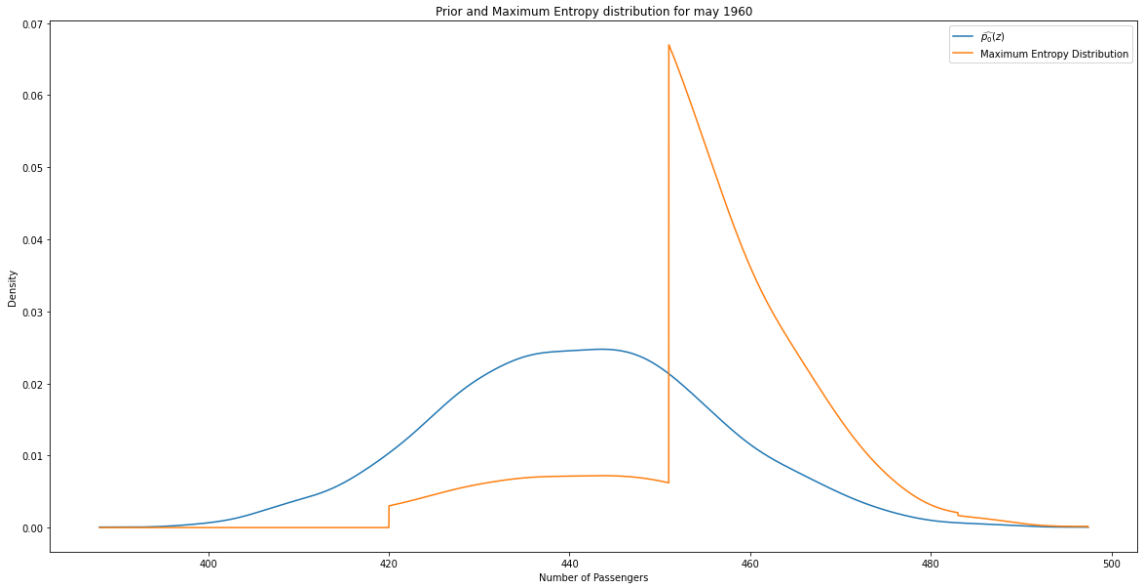
$$\text{This however means nothing more than: } E[F_k] = f_k, \qquad k = 1, ..., m$$

We are finally here: we have to find λ, so that the expected values of the functions F_k match the given constraints.

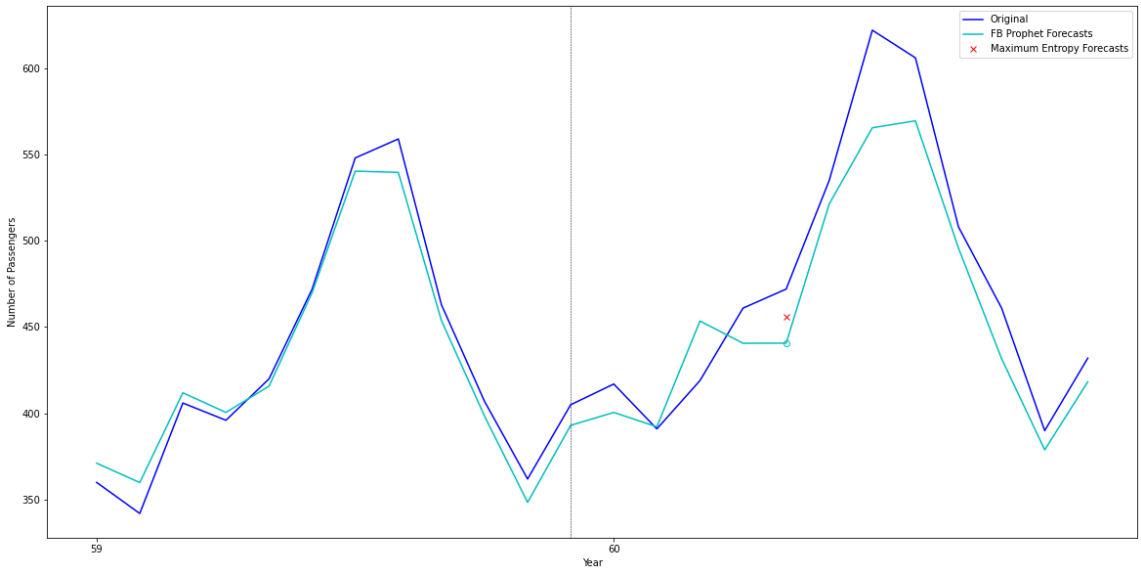
2.2 Implementation Issue

As the number of constraints rises, the numerical solution to the system of equations becomes increasingly harder to find. Due to the problem of multiple local minima, we refrain from using a gradient-based algorithm and instead use a heuristic algorithm. In our case, it is the particle swarm algorithm (Python package pyswarm).

The result of applying the maximum entropy distribution can be summarized with the picture below and we can visibly see the improvement on the prediction after including the expert's opinion in the form of constraints.



Using this maximum entropy distribution to get the prediction corrects the predicted value to the right direction when expert's opinion is valid.



Reference:

1. https://blog.codecentric.de/en/2019/02/forecasts-machine-learning-facebook-prophet-maximum-entropy/