Assignment 2: Call Center Modeling

Ash

2018/12/06

1 Call Center Data Modeling

Loading and preparing the data

```
#Importing relevant libraries
% matplotlib notebook
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import matplotlib
import pandas as pd

#Loading and print basic infos about the data
data = np.loadtxt('/Users/ash/Downloads/call_center.csv')
print('Size of data set:', len(data))
print('First 10 values in data set:', data[:10])
print('Sum of data set:', sum(data))

# Split the data into 24 separate series, one for each hour of the day
# Code from Scheffler's gist
# (https://gist.github.com/cscheffler/6a03c9473297f21b78363ec7301d19d8#file-cs146-2-2-pre-class-work-ipynb)
current_time = 0
waiting_times_per_hour = [[] for _ in range(24)]  # Make 24 empty lists, one per hour
for t in data:
    current_hour = int(current_time // 60)
    current_time += t
    waiting_times_per_hour[current_hour].append(t)
```

--> Output:

```
Size of data set: 5891
First 10 values in data set: [ 5.36  2.48  8.08  1.54 11.1  10.7  21.
   11.1  14.7  32.9 ]
Sum of data set: 1442.145437310004
```

Modeling call rates

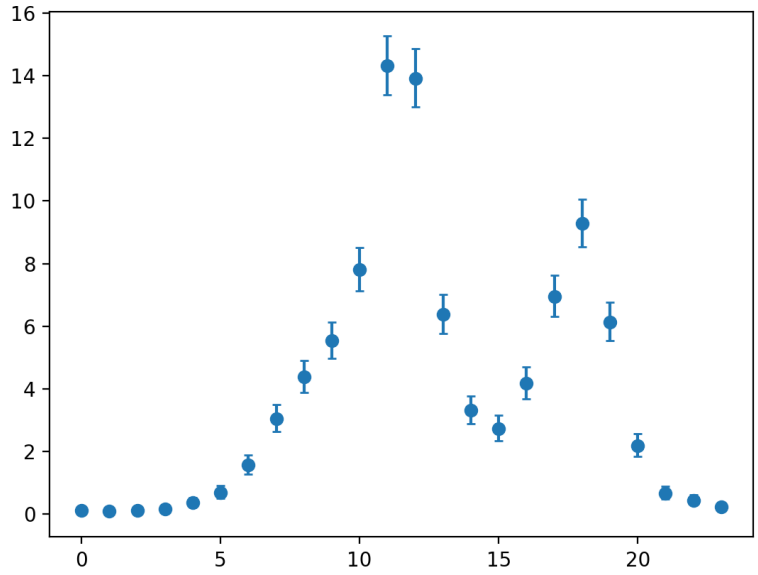

Figure 1: Call Rates For 24 Hour With 95% Confidence Intervals

```
# Modeling the prior as a gamma distribution and the likelihood as an
    exponential distribution
# with the following parameters
a_prior = 1
b_prior = 2.5
rate_likelihood = 3

# Defining a function to update the posterior with available data based on
    the conjugate prior formula
def define_post(data, aprior, bprior):
    n = len(data)
    s = sum(data)
    a_post, b_post = aprior+n, bprior+s
    posterior = stats.gamma(a = a_post, scale = 1/b_post)
    return posterior

# Defining the prior and the likelihood
prior = stats.gamma(a = a_prior, scale = 1/b_prior)
likelihood = stats.expon(scale=1/rate_likelihood)

# Defining lists and arrays to record the means and error bars
mins = np.ones(24)
maxs = np.ones(24)
means = []

# Looping over all 24 hours and using the defined function to calculate the
```

	95% CI Lower Bound	95% Upper Bound	Mean
0	0.055032	0.229807	0.127470
1	0.031400	0.166398	0.085564
2	0.056987	0.218279	0.124626
3	0.073646	0.307533	0.170583
4	0.225710	0.525518	0.360160
5	0.498491	0.909411	0.688805
6	1.265881	1.886619	1.561019
7	2.626194	3.491338	3.043599
8	3.879804	4.915417	4.382520
9	4.967570	6.135692	5.536435
10	7.127521	8.512193	7.804711
11	13.399793	15.278563	14.323986
12	13.007577	14.856947	13.917112
13	5.774776	7.022556	6.383628
14	2.876432	3.778503	3.312317
15	2.338408	3.159930	2.733944
16	3.682471	4.692592	4.172452
17	6.305547	7.616780	6.945902
18	8.539682	10.049862	9.279620
19	5.540454	6.768249	6.139212
20	1.840630	2.574385	2.192358
21	0.491912	0.879505	0.671896
22	0.296552	0.617905	0.442803
23	0.124411	0.376749	0.233653

Figure 2: Call Rates For 24 Hour With 95% Confidence Intervals - Table

```
    posterior, which
# then tell us the means and 95% confidence intervals
for _ in range(24):
    posterior = define_post(waiting_times_per_hour[_], a_prior, b_prior)
    min_, max_ = posterior.interval(0.95)
    mean_ = posterior.mean()
    means.append(mean_)
    mins[_] = min_
    maxs[_] = max_

# Plotting the means and the error bars
intervals = [np.absolute(mins-means), np.absolute(maxs-means)]
plt.errorbar(x=[_ for _ in range(24)], y=means, yerr= intervals,fmt='o',
    capsize=2)
plt.show()

# Constructing table
d = {'Mean': means, '95% CI Lower Bound': mins, '95% Upper Bound': maxs}
df = pd.DataFrame(data=d)
print(df)
```

−− >Output: fig.1 and fig.2

Report to clients

You can expect (with 95% certainty) the most call per minute from 11AM-12AM and from 12AM-1PM, at around 13 call per minute at least and 15 call per minute at most, so you should plan your agents accordingly. The least busy time is from 9PM-5AM, when there is less than 1 call per minute, and this is also a period in which there is less variability in the number of call per minute as you can see in the graph. After 5AM the number of call will increase quickly, then dip down at around 3PM. After that you can expect another slight peak at 6PM, then things will quiet down for the night. Note that during peak hours of the day (11AM-1PM) you will probably see the call rate will vary more than the other hours, so just because there is only 12 call per minute at between 11AM-12PM today does not mean that tomorrow the call rate in that period cannot reach 15 call per minute. Therefore you should always arrange slightly more agents than needed in this period, while you can assign a constant number of agents in the night hours where the variance is not that much.

2 Stretch Goals

Normal Likelihood Re-parameterization

We have the normal distribution pdf:

$$\mathcal{N}_{\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Since in the pdf of the normal distribution σ^2 is a parameter we can simply substitute $\sigma = \tau^{-1/2}$ in to the pdf to re-parameterize the function:

$$\begin{aligned}\mathcal{N}_{\tau} &= \frac{1}{\sqrt{2\pi(\tau^{-1/2})^2}} e^{-\frac{(x-\mu)^2}{2(\tau^{-1/2})^2}} \\ &= \frac{\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{\tau(x-\mu)^2}{2}}\end{aligned}$$

Normal-Inverse-Gamma Change Of Variable

Since in the normal-inverse-gamma (NIG) pdf function σ^2 is not a parameter but rather a variable, we need to derive the change of variable by using the cdf:

$$\begin{aligned}pdf_{NIG}(x, \sigma^2) = pdf_{x,\sigma^2} &= \sqrt{\frac{\lambda}{2\pi\sigma^2}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} e^{-\frac{2\beta+\lambda(x-\mu)^2}{2\sigma^2}} \\ &= \sqrt{\frac{\lambda}{2\pi}} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+3/2} e^{-\frac{2\beta+\lambda(x-\mu)^2}{2\sigma^2}}\end{aligned}$$

With $\sigma = \tau^{-1/2}$ we have:

$$\begin{aligned} cdf_{x,\tau}(\tau_0) = P(\tau \leq \tau_0) &= P\left(\frac{1}{\sigma^2} \leq \tau_0\right) \\ &= P\left(\sigma^2 \geq \tau_0^{-1}\right) \\ &= 1 - cdf_{x,\tau}(\tau_0^{-1})) \end{aligned}$$

So

$$\begin{aligned} pdf_{x,\tau} &= \frac{\partial cdf_{x,\tau}(\tau_0)}{\partial \tau} \\ &= \frac{\partial\left[1 - cdf_{x,\tau}(\tau_0^{-1}))\right]}{\partial \tau} \\ &= -\frac{\partial\left[cdf_{x,\tau}(\tau_0^{-1}))\right]}{\partial \tau} \\ &= -\frac{\partial\left[cdf_{x,\tau}(\tau_0^{-1}))\right]}{\partial \tau}\frac{\partial \sigma^2}{\partial \sigma^2} \\ &= -\frac{\partial\left[cdf_{x,\sigma^2}(\tau_0^{-1}))\right]}{\partial \sigma^2}\frac{\partial \sigma^2}{\partial \tau} \\ &= -pdf_{x,\sigma^2}(\tau^{-1}))\frac{d\sigma^2}{d\tau} \\ &= -\sqrt{\frac{\lambda}{2\pi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\tau^{-1}}\right)^{\alpha+3/2} e^{-\frac{2\beta+\lambda(x-\mu)^2}{2\tau^{-1}}}\frac{d\left(\tau^{-1}\right)}{d\tau} \\ &= -\sqrt{\frac{\lambda}{2\pi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha+3/2}e^{-\frac{2\tau\beta+\tau\lambda(x-\mu)^2}{2}}\left(-\tau^{-2}\right) \\ &= \sqrt{\frac{\lambda}{2\pi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1/2}e^{-\frac{2\tau\beta+\tau\lambda(x-\mu)^2}{2}} \\ &= \sqrt{\frac{\lambda}{2\pi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1/2}e^{-\tau\beta}e^{-\frac{\tau\lambda(x-\mu)^2}{2}} \end{aligned}$$

which is the form of the normal-gamma distribution. In the above transformation we switched τ_0 back to τ because τ_0 is only a placeholder and is interchangable to τ.

In the above transformation we have to use the chain rule of derivation because we are trying to change variable for the probability function, and in order to differentiate the cdf into the pdf of the new variable we need to multiply it by the derivative of the old variable over the new one. If we choose to use the formula to change variable rather than derive them from scratch as the above we will have to multiply by the absolute value of the derivative of the old variable over the new one since we are changing from the variable σ which has the domain of the entire real line to the variable $\tau = 1/\sigma^2$ which has the domain of the positive real line, we need to make sure the derivative is

valid for the negative real line as well, which we can do by taking the absolute value: $\left|\frac{d\sigma}{d\tau}\right|$.