Chris Piech
CS109

Section #7
Nov 14, 2018

Section #7: Maximum Likelihood Honor Code

1. **Single Match**:

   Let A_i be the event that decision point i is matched. We note that a match occurs when both students make the more popular choice or when both students make the less popular choice. $P(A_i) = P(\text{Both more popular}) + P(\text{Both less popular}) = p^2 + (1-p)^2$.

   Let M be a random variable for the number of matches. It is easy to see that each of the 1000 decisions is an independent Bernoulli experiment with probability of success $p' = p^2+(1-p)^2$. Therefore $M \sim Bin(1000, p')$.

   We can use a Normal distribution to approximate a binomial. We approximate $M \sim Bin(1000, p')$ with Normal random variable $Y \sim N(1000p', 1000(1-p')p')$.

2. **Maximum Match**:

   For this problem, we use Maximum Likelihood Estimator (MLE) to estimate the parameters $\theta = (\mu, \beta)$.

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} f(Y^{(i)} = y^{(i)} \mid \theta) \\
LL(\theta) &= \log \prod_{i=1}^{n} f(Y^{(i)} = y^{(i)} \mid \theta) \\
&= \sum_{i=1}^{n} \log f(Y^{(i)} = y^{(i)} \mid \theta) \\
&= \sum_{i=1}^{n} \log \frac{1}{\beta} e^{-(z_i + e^{-z_i})} \qquad \text{where } z_i = \frac{y^{(i)} - \mu}{\beta} \\
&= \sum_{i=1}^{n} \log \frac{1}{\beta} + \sum_{i=1}^{n} -(z_i + e^{-z_i}) \\
&= -n \log(\beta) + \sum_{i=1}^{n} -(z_i + e^{-z_i})
\end{aligned}
$$

   Now we must choose the values of $\theta = (\mu, \beta)$ that maximize our log-likelihood function. To solve this argmax, we will use Gradient Ascent. First, we need to find the first derivative of the log-likelihood function with respect to our parameters.

$$
\begin{aligned}
\frac{\partial LL(\theta)}{\partial \mu} &= \frac{\partial}{\partial \mu}\Big[-n\log(\beta) + \sum_{i=1}^{n} -(z_i + e^{-z_i})\Big] \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \mu}\Big[-(z_i + e^{-z_i})\Big] \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial z_i}\Big[-(z_i + e^{-z_i})\Big]\frac{\partial z_i}{\partial \mu} && \text{By the Chain Rule} \\
&= \sum_{i=1}^{n} \Big[-1 + e^{-z_i}\Big]\Big[-\frac{1}{\beta}\Big] \\
&= \frac{1}{\beta}\sum_{i=1}^{n} 1 - e^{-z_i}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial LL(\theta)}{\partial \beta} &= \frac{\partial}{\partial \beta}\Big[-n\log(\beta) + \sum_{i=1}^{n} -(z_i + e^{-z_i})\Big] \\
&= -\frac{n}{\beta} + \sum_{i=1}^{n} \frac{\partial}{\partial \beta}\Big[-(z_i + e^{-z_i})\Big] \\
&= -\frac{n}{\beta} + \sum_{i=1}^{n} \frac{\partial}{\partial z_i}\Big[-(z_i + e^{-z_i})\Big]\frac{\partial z_i}{\partial \beta} && \text{By the Chain Rule} \\
&= -\frac{n}{\beta} + \sum_{i=1}^{n} \Big[-1 + e^{-z_i}\Big]\Big[\frac{\mu - y^{(i)}}{\beta^2}\Big] && \text{Where the last term equals } \frac{\partial z_i}{\partial \beta}
\end{aligned}
$$

Now that we know the derivative of the log-likelihood function with respect to each parameter, we have the information we would need to perform gradient ascent. We would initialize our values of θ, either to zero or to random values, and then iteratively take a small step in the direction of the gradient for each variable in θ (μ and β) and recalculate the gradient until the gradient approaches zero.

3. Understanding:

$P(Y >= 90) = 0.00000017180200395650047$, or nearly 1 in 6 million.

```
from scipy.stats import gumbel_r
print(1 - gumbel_r.cdf(90, 9, 5.2))
```