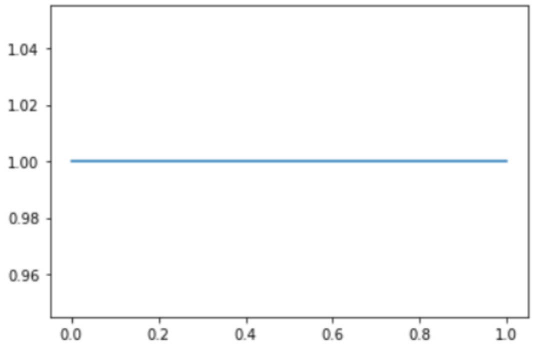
```
In [3]: a=1;b=1 #beta(1,1)
        hypos=np.linspace(0,1,101)
        x=beta.pdf(hypos,a,b)
        prior=Pmf(x,hypos)
```

```
In [4]: plt.plot(hypos,x) #prior plot
Out[4]: [<matplotlib.lines.Line2D at 0x25d2cb3af70>]
```



- Prob (Yes)

$Prob(tail)\,Prob(Yes|tail) + Prob(head)\,Prob(Yes|head)$

$= 0.5 \cdot hypo + 0.5 + 1 = \frac{1}{2}(hypo+1)$

- Prob (No)

$Prob(tail)\,Prob(No|tail) + Prob(head)\,Prob(No|head)$

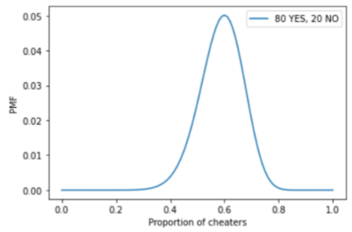
$= 0.5 \cdot (1-hypo) + 0.5 + 0 = \frac{1}{2}(1-hypo)$

```
In [5]: likelihood={'Y':0.5+hypos/2, 'N':(1-hypos)/2} #Q1 WHY?? IS THIS BINOMIAL MODEL?
```

Q1 : Likelihood 함수가 이와 같이 나오는 이유를 설명해주세요. 그리고 이와 같은 sampling model은 binomial model인가요? 왜 그런가요?

```
In [6]: dataset='Y'*80+'N'*20 #DATA 80 YES, 20 NO
        posterior1=prior.copy()
        for data in dataset:
            posterior1 *= likelihood[data]
        posterior1.normalize()
Out[6]: 3.694513913396701e-21
```

```
In [7]: posterior1.plot(label='80 YES, 20 NO',xlabel='Proportion of cheaters',ylabel='PMF')
        plt.legend() #Q2 WHY MODE ON 0.6? DATA IS 80 YES & 20 NO. WHY NOT 0.8?
Out[7]: <matplotlib.legend.Legend at 0x25d2ad13a90>
```



- $P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$ $(P(\theta)=1)$

$= C(Y)\,\theta^{a+Y-1}(1-\theta)^{b+n-Y-1}$ $\left(C(Y) = \frac{\Gamma(a+b+n)}{\Gamma(a+Y)\Gamma(b+n-Y)}\right)$

$= beta(a+Y,\ b+n-Y)$

- $E[\theta|Y] = \frac{a+Y}{a+b+n}$

$= \frac{a+b}{a+b+n} \cdot \underbrace{\frac{a}{a+b}}_{\text{Prior exp}} + \frac{n}{a+b+n} \cdot \underbrace{\frac{Y}{n}}_{\text{data avg}}$

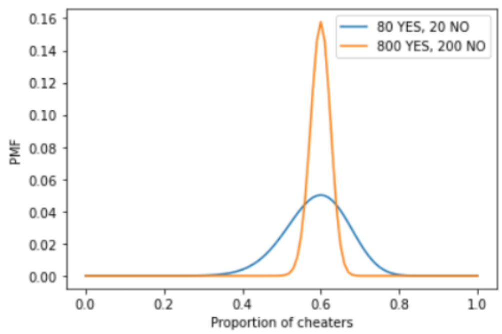
Q2 : 우리가 구한 Posterior 분포의 꼴이 왜 다음과 같이 나왔을까요? prior의 꼴과 data와 관련하여 설명해주세요. 그리고 data는 YES와 NO의 비율이 4:1인데 왜 posterior의 mode는 0.8이 아니라 0.6인걸까요?

⇒ 적용 : $(a=b=1,\ n=100,\ Y=80)$ ⇒ $beta(81, 21)$, $E[\theta|Y] = \frac{1+80}{1+1+100} = \frac{81}{102} \approx 0.8$

⊛ likelihood 설정시 hypo를 그대로 사용하지 않음.

$E[\theta|Y] = E\left[\frac{1}{2}(1+hypo)\,|\,Y\right] = 0.8 \Rightarrow E[hypo|Y] = 0.6$

```
<matplotlib.legend.Legend at 0x25d2d3c0760>
```



① $n=100,\ Y=80$

⇒ $beta(101, 21)$

② $n=1000,\ Y=800$

⇒ $beta(1001, 201)$

Q3: 두 posterior 분포는 왜 이렇게 다른 꼴이 나오게 되었을까요?

Homework

2. Assume home runs per game at Citizens Bank Park follow a Poisson distribution with parameter θ. Assume for θ a Gamma prior distribution with shape parameter $\alpha = 4$ and rate parameter $\beta = 2$.

1) Write an expression for the prior density $\phi(\theta)$. Find the prior mean and prior SD.

2) Suppose a single game with 1 home run is observed. Write the likelihood function.

3) Write an expression for the posterior distribution of θ given a single game with 1 home run. Find the posterior mean and posterior SD.

1) $P(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-b\theta} = \frac{2^4}{\Gamma(4)} \theta^3 e^{-2\theta}$ $\quad E[\theta] = \frac{\alpha}{\beta} = \frac{4}{2} = 2$ $\quad Var[\theta] = \frac{\alpha}{\beta^2} = 1$

$sd[\theta] = 1$

2) $P(y_i|\theta) = \frac{\theta^{y_i} e^{-\theta}}{y_i!}$ $\rightarrow y_1 = 1$, $P(y_1 = 1|\theta) = \theta \cdot e^{-\theta}$

3) $P(\theta|y_1, \cdots, y_n) = \frac{P(\theta)P(y_1, \cdots, y_n|\theta)}{P(y_1, \cdots, y_n)} = C \cdot \theta^{a+\Sigma y_i - 1} e^{-(n+b)\theta}$ $\sim gamma(a+\Sigma y_i, n+b)$

$\Longrightarrow$ 적용 $(n=1, y_1=1)$ $\Rightarrow gamma(5, 3)$ $\Rightarrow \begin{cases} mean: \frac{\alpha}{\beta} = \frac{5}{3} \\ Var: \frac{\alpha}{\beta^2} = \frac{5}{3^2} \end{cases}$ $\quad sd: \frac{\sqrt{5}}{3}$