An Exploration of Bayesian Inference: Coin Toss

Before any experiments lets assume that the coin is fair via a weak prior PDF:

$$P(\theta) = \frac{1}{Beta(\alpha,\beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$P(\theta) = \frac{1}{Beta(5,5)} \cdot \theta^{4}(1-\theta)^{4} = 630 \cdot \theta^{4}(1-\theta)^{4}$$

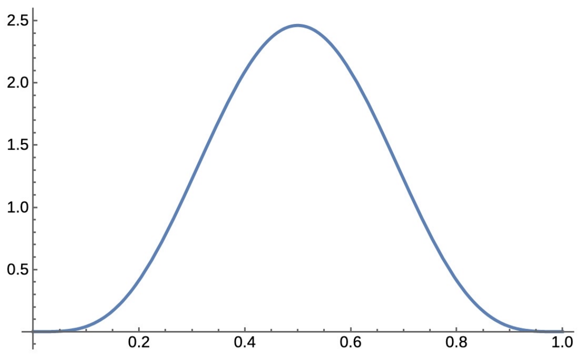
lets check this out in Mathematica:

```
In[70]:= (* Define the PDF *)
         P[x_] := (1/Beta[5, 5]) * x^4 * (1 - x)^4;
         TraditionalForm[P[θ]]
Out[71]//TraditionalForm=
         630 (1 − θ)^4 θ^4

In[72]:= (* Validate the PDF integrates to 1 *)
         Integrate[P[x], {x, 0, 1}] == 1
Out[72]= True

In[73]:= (* Plot the PDF *)
         Plot[P[x], {x, 0, 1}]
```

Out[73]=



lets say we perform 10 coin flips (trials) and observe 7 heads

next, we want to combine our prior belief and observed data to form our posterior belief. To do this we can leverage Bayes' theorem:

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{P(D)}$$

in the coin flip case, we find that:

$$P(D|\theta) = L(\theta|D) = L(\theta|n,y)$$
$$= \binom{n}{y}\theta^{y}(1-\theta)^{n-y} = \binom{10}{7}\theta^{7}(1-\theta)^{3}$$

lets now take a look at the marginal likelihood or evidence term $P(D)$ in the denominator. By definition:

$$P(D) = \int_0^1 P(D|\theta)\cdot P(\theta)\,d\theta$$

substituting in what we know:

$$P(D) = \int_0^1 \left(\binom{10}{7}\theta^{7}(1-\theta)^{3}\right)\left(\frac{1}{Beta(5,5)}\theta^{4}(1-\theta)^{4}\right)d\theta$$
$$= \binom{10}{7}\left(\frac{1}{Beta(5,5)}\right)\int_0^1 \theta^{11}\cdot(1-\theta)^{7}\,d\theta$$

recall that $Beta(\alpha,\beta) = \int t^{\alpha-1}(1-t)^{\beta-1}\,dt$

so, $P(D) = \binom{10}{7}\dfrac{Beta(12,8)}{Beta(5,5)}$

using all of our new knowledge, lets take a look back at:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{\binom{10}{7}\frac{1}{Beta(5,5)}\theta^{11}(1-\theta)^{7}}{\binom{10}{7}\frac{Beta(12,8)}{Beta(5,5)}}$$

simplifying this we get a clean, normalized posterior PDF:

$$P(\theta|D) = \frac{1}{Beta(12,8)} \theta^{11}(1-\theta)^{7}$$

note that we started w/ a very similar looking prior belief PDF of:

$$P(\theta) = \frac{1}{Beta(5,5)} \cdot \theta^{4}(1-\theta)^{4}$$

now, more generally we can write the posterior PDF for our coin toss example as:

$$P(\theta|D) = \frac{1}{Beta(\alpha_{prior}+k, \beta_{prior}+(n-k))} \theta^{\alpha_{prior}-k}(1-\theta)^{\beta-(n-k)}$$

lets check this out in Mathematica:

In[107]:=
```
(* Defining the posterior belief PDF *)
PosteriorP[α_, β_, n_, k_, θ_] := (1/Beta[α+k, β+(n-k)]) * θ^((α-1)+k) * (1-θ)^((β-1)+(n-k));
TraditionalForm[PosteriorP[α, β, n, k, θ]]
```

Out[108]//TraditionalForm=

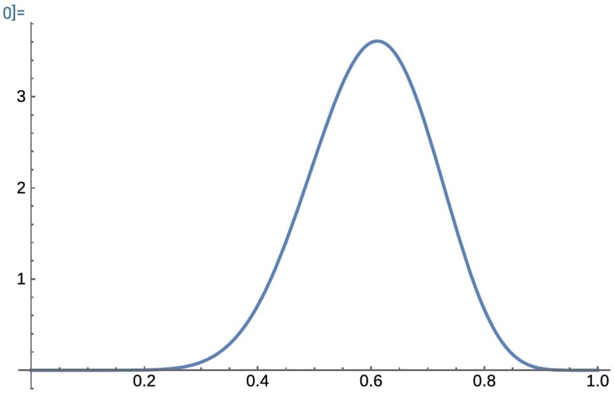
$$\frac{\theta^{\alpha+k-1}(1-\theta)^{\beta-k+n-1}}{\mathrm{B}(k+\alpha, -k+n+\beta)}$$

In[109]:=
```
(* Validate the PDF integrates to 1 *)
Integrate[PosteriorP[5, 5, 10, 7, θ], {θ, 0, 1}] == 1
```

Out[109]=
True

In[110]:=
```
(* Plot the posterior belief PDF *)
Plot[PosteriorP[5, 5, 10, 7, θ], {θ, 0, 1}]
```

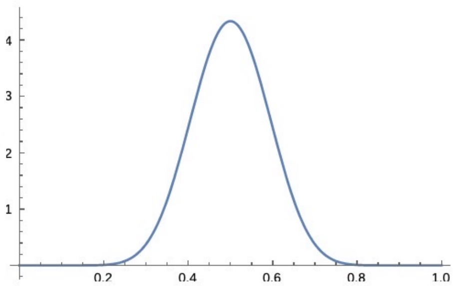
Out[110]=



what happens if we do more trials w/ a fair coin?

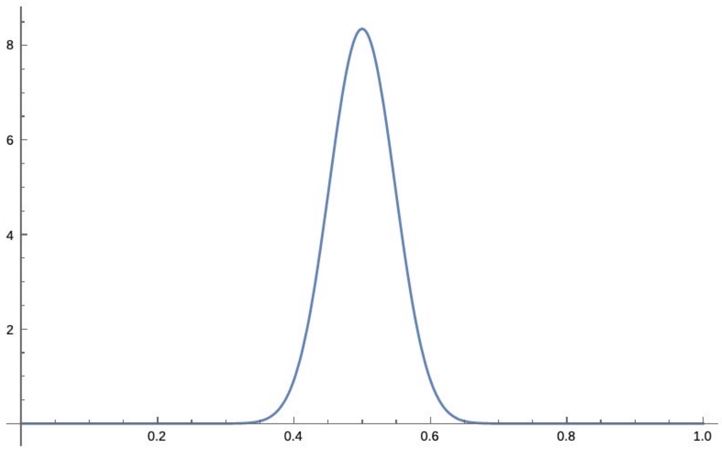
```
In[117]:= (* Now what happens if we do increasingly more trials with a fair coin? *)
Plot[PosteriorP[5, 5, 20, 10, θ], {θ, 0, 1}]
```

Out[117]=



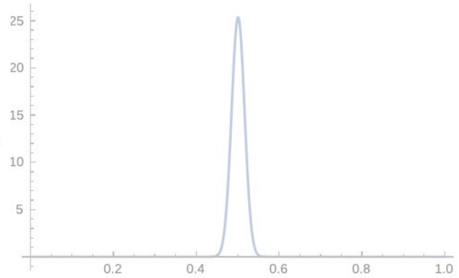

```
In[130]:= (* Even more trials? *)
Plot[PosteriorP[5, 5, 100, 50, θ], {θ, 0, 1}, PlotRange → All]
```

Out[130]=




```
(* Even more trials? *)
Plot[PosteriorP[5, 5, 1000, 500, θ], {θ, 0, 1}, PlotRange → All]
General: 0.0000204286^505 is too small to represent as a normalized machine number; precision may be lost.
```

Out[132]=



well, lets note that the variance of the Beta distribution is given by:

$$Var(\theta) = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

therefore, the variance of our posterior distribution is given by:

$$PosteriorVar(\theta) = \frac{(\alpha_{prior}+K)(\beta_{prior}+(n-K))}{((\alpha_{prior}+K)+(\beta_{prior}+(n-K)))^2((\alpha_{prior}+K)+(\beta_{prior}+(n-K))+1)}$$

if coin flip process was fair (50/50) and we perform an increasing number of trials, the Law of Large numbers tells us that given a sample of independent and identically distributed (i.i.d.) values the sample mean converges to the true mean (expected value). Here, the expected value is:

let $x=1$ be heads and $x=0$ be tails

$$E[X] = 1\cdot\left(\frac{1}{2}\right) + 0\cdot\left(\frac{1}{2}\right) = \frac{1}{2}$$

therefore, when flipping the coin n times, the expected number of heads K converges to:

$$E[K] = n \cdot E[X] = \frac{n}{2}$$

therefore:

$$\lim_{n\to\infty} \frac{\left(\alpha_{prior}+\frac{n}{2}\right)\left(\beta_{prior}+\left(n-\frac{n}{2}\right)\right)}{\left(\left(\alpha_{prior}+\frac{n}{2}\right)+\left(\beta_{prior}+\left(n-\frac{n}{2}\right)\right)\right)^2\left(\left(\alpha_{prior}+\frac{n}{2}\right)+\left(\beta_{prior}+\left(n-\frac{n}{2}\right)\right)+1\right)} = 0$$

which means as we perform an increasing number of coin flips, the variance of our PDF of the parameter θ will tend to 0.