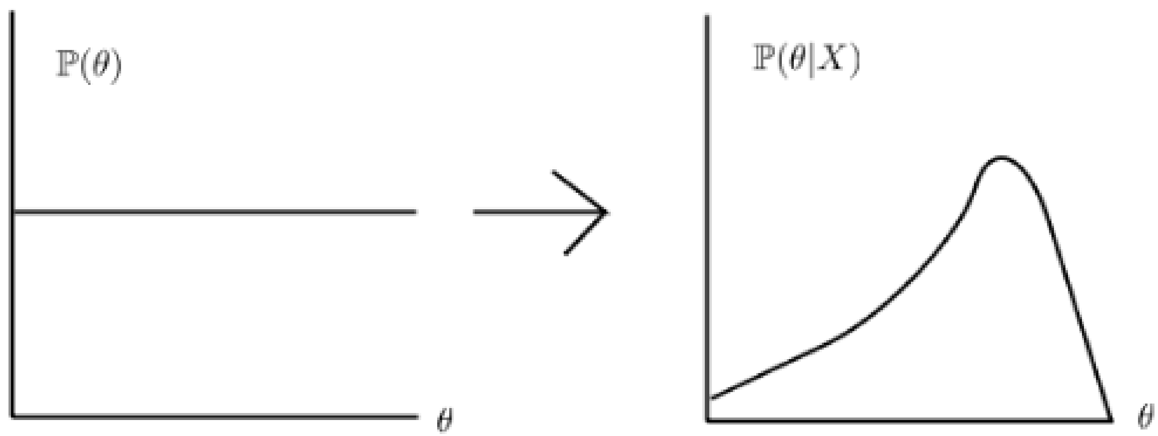
We let $\mathcal{F}$ be Bernoulli with $X = \langle 0, 1, 1 \rangle$ and $\theta \sim U(0,1)$. This means that we give equal weightage to all values for θ in between 0 and 1. If $\mathbb{P}(\theta \mid X) = 12\theta^2(1-\theta)$, then we went from $\mathbb{P}(\theta)$, the prior distribution, to $\mathbb{P}(\theta \mid X)$, the posterior distribution, or,



This shows a skewness towards 1 because $\hat{\theta}_{\text{MAP}} = \frac{2}{3} = \hat{\theta}_{\text{MLE}}$.

Note: Under the principle of indifference,

$$\hat{\theta}_{\text{MAP}} = \hat{\theta}_{\text{MLE}}$$

Let $\mathcal{F}$ be Bernoulli with $X = \langle 0, 1, 1 \rangle$ and $\theta \sim U(0,1)$. Then

$$\overbrace{\mathbb{P}(\theta \mid X)}^{\text{all data}} = \frac{\mathbb{P}(X \mid \theta)\,\mathbb{P}(\theta)}{\mathbb{P}(X)} = \frac{\mathbb{P}(X \mid \theta)\,\mathbb{P}(\theta)}{\int_{\Theta_0} \mathbb{P}(X \mid \theta)\,\mathbb{P}(\theta)\; d\theta}$$

where $\mathbb{P}(\theta) = 1$. Then, for this model,

$$\begin{aligned}
\mathbb{P}(X \mid \theta) &= \prod_{i=1}^{n} \mathbb{P}(x_i \mid \theta) \\
&= \prod_{i=1}^{n} \theta^{x_1}(1-\theta)^{1-x_i} \\
&= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\
&= \theta^{x}(1-\theta)^{n-x} \text{ where } x = \sum_{i}^{n} x_i
\end{aligned}$$

Plugging this back into $\mathbb{P}(\theta \mid X)$ gives:

$$\mathbb{P}(\theta \mid X) = \frac{\theta^x(1-\theta)^{n-x}}{\int_0^1 \theta^x(1-\theta)^{n-x}\, d\theta}$$

which can only be computed numerically.

Definition 0.1. Beta Function:

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

Using the beta function, we get

$$\mathbb{P}\left(\theta \mid X\right) = \frac{\theta^x(1-\theta)^{n-x}}{\mathrm{B}(x+1, n-x+1)}$$

Let's look at the random variable $X \sim \mathrm{Beta}(\alpha, \beta)$ and its distribution.

$$X \sim \mathrm{Beta}(\alpha, \beta) := \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

Its support is $(0, 1)$.

If $f(x)$ is a pdf, then $\int_{\mathrm{Supp}[X]} f(x)\,dx = 1$. Using this information, show that $\mathrm{Beta}(\alpha, \beta)$ is a pdf.

$$\int_0^1 \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{1}{\mathrm{B}(\alpha, \beta)} \overbrace{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx}^{\mathrm{B}(\alpha,\beta)} = 1\checkmark$$

Its parameter space is $\alpha > 0$ and $\beta > 0$ where its finite.

Definition 0.2. Gamma Function:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1}\,dt$$

which can only be computed numerically.

Properties of the Gamma Function:

1. $\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
2. $\Gamma(x) = (x-1)!$ where $x \in \mathbb{N}$
3. $\Gamma(x) = (x-1)\Gamma(x-1)$ valid $\forall x$
4. $\Gamma(x+1) = x\Gamma(x)$

What's the expected value of a Beta distribution?

$$\begin{aligned}
\mathrm{E}[X] &= \int_{\Theta_0} x f(x)\,dx \\
&= \int_0^1 x \cdot \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx \\
&= \frac{\mathrm{B}(\alpha+1,\beta)}{\mathrm{B}(\alpha,\beta)} \\
&= \frac{[\Gamma(\alpha+1)\Gamma(\beta)]/[\Gamma(\alpha+\beta+1)]}{[\Gamma(\alpha)\Gamma(\beta)]/[\Gamma(\alpha+\beta)]} \\
&= \frac{\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

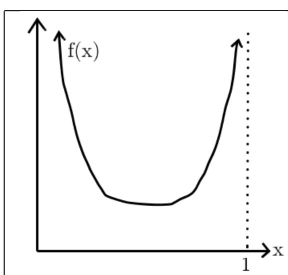
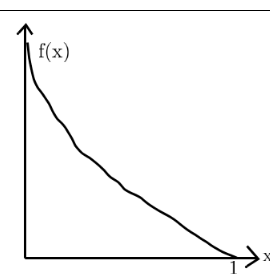
What's the mode of X if X is Beta?

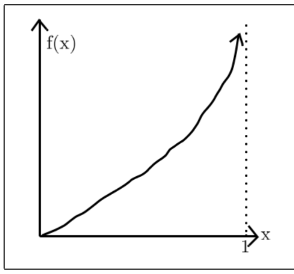
$$\begin{aligned}
\mathrm{Mode}[X] &= \underset{x\in\mathrm{Supp}[X]}{\mathrm{argmax}}\{f(x)\} \\
&= \mathrm{argmax}\{\frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\} \\
&= \mathrm{argmax}\{x^{\alpha-1}(1-x)^{\beta-1}\} \\
&= \mathrm{argmax}\{(\alpha-1)\ln(x) + (\beta-1)\ln(1-x)\}
\end{aligned}$$

If we differentiate this function and set it equal to 0, we will find x.

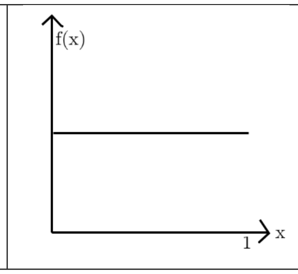
$$\begin{aligned}
\frac{d}{dx}\Big[(\alpha-1)\ln(x) + (\beta-1)\ln(1-x)\Big] &= \frac{\alpha-1}{x} - \frac{\beta-1}{1-x} = 0 \\
x &= \frac{\alpha-1}{\alpha+\beta-2} \text{ only for } \alpha>1, \beta>1
\end{aligned}$$

Different Types of Gamma Distributions

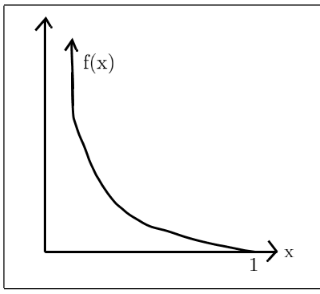
	$\alpha=\beta=0.5$		$\alpha=1, \beta>1$



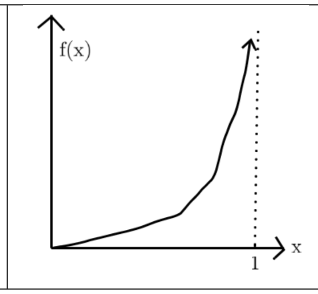
$\alpha > 1, \beta = 1$



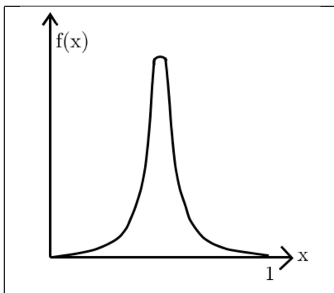
$\alpha = \beta = 1$



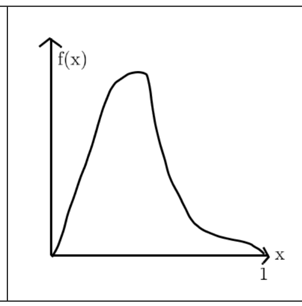
$\alpha < 1, \beta > 1$



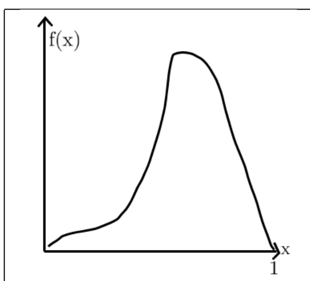
$\alpha > 1, \beta < 1$



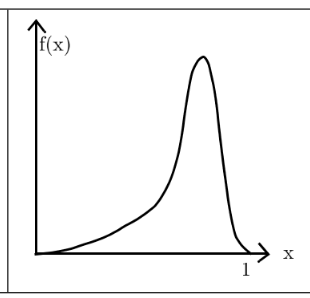
$\alpha = \beta = 100$



$\alpha = 1.01, \beta >> 1$



$\alpha >> 1, \beta = 1.01$



$\alpha = 100, \beta = 10$

Let's say $\mathcal{F}$ is Binomial with n known and $\theta \sim U(0,1) = \text{Beta}(1,1)$. Refresher: $\text{Binom}(n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$. Then:

$$\begin{aligned}
\mathbb{P}(\theta \mid X) &= \frac{\mathbb{P}(X \mid \theta)\overbrace{\mathbb{P}(\theta)}^{1}}{\underbrace{\mathbb{P}(X)}_{\int_{\Theta_0}\mathbb{P}(X \mid \theta)\,d\theta}} \\
&= \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}}{\int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\,d\theta} \\
&= \text{Beta}(x+1, n-x+1)
\end{aligned}$$

Before we transformed $\mathbb{P}(\theta) \to \mathbb{P}(\theta \mid X)$ using X (the data). Here we transformed $\text{Beta}(1,1) \to \text{Beta}(x+1, n-x+1)$ where the first value is α and the second is β. For example, if $n = 10$ and $x = 7$, then $\theta|X \sim \text{Beta}(8,4)$. What's $\hat{\theta}_{\text{MLE}}$?

$$\hat{\theta}_{\text{MLE}} = \hat{\theta}_{\text{MAP}} = \text{Mode}[\theta|X] = \frac{\alpha-1}{\alpha+\beta-1} = \frac{7}{10} = 0.7$$

Definition 0.3. Minimum Mean Square Error:

$$\hat{\theta}_{\text{MMSE}} := \text{E}[\theta|X]$$

where E is the posterior mean or expectation.

What's $\hat{\theta}_{\text{MMSE}}$ of the above distribution?

$$\hat{\theta}_{\text{MMSE}} = \text{E}[\theta|X] = \frac{\alpha}{\alpha+\beta} = \frac{2}{3} = 0.67$$

Definition 0.4. Mean Absolute Error:

$$\hat{\theta}_{\text{MAE}} = \text{Med}[\theta|X]$$

where Med is the posterior median.

Note: MAE can only be computed numerically using a computer. If using R, the command is: qbeta(0.5, α, β).
In this distribution, $\hat{\theta}_{\text{MAE}}$ comes out to be 0.676.

Definition 0.5. Quantile: If X is a continuous random variable,

$$\text{Quantile}[X,p] = F^{-1}(p)$$

Thus we say that $\text{Med}[X] = \text{Quantile}[X,\ 0.5] = F^{-1}(\frac{1}{2})$.

Let say $\mathcal{F}$ is Binomial and $\theta \sim \text{Beta}(\alpha, \beta)$ with appropriately chosen α and β. Then:

$$\begin{aligned}
\mathbb{P}(\theta \mid X) &= \frac{\mathbb{P}(X \mid \theta)\,\mathbb{P}(\theta)}{\mathbb{P}(X)} \\
&= \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x} \cdot \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 \binom{n}{x}\theta^x(1-\theta)^{n-x}\frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}\,d\theta} \\
&= \frac{\theta^{x-\alpha-1}(1-\theta)^{n-x+\beta-1}}{\int_0^1 \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}\,d\theta} \\
&= \frac{1}{\mathrm{B}(x+\alpha, n-x+\beta)}\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \\
&= \text{Beta}(x+\alpha, n-x+\beta)
\end{aligned}$$

Here we have went from Beta to Beta using X. We call this conjugacy, where the prior and posterior are of the same family. In other words, the beta is conjugate prior for the binomial model.