EXAMPLE: COIN TOSSED

\[
\begin{align*}
&\text{Suppose } \theta \sim \text{Beta}(\alpha, \beta) \implies \theta | x \sim \text{Beta}(\alpha + x, \beta + n - x) \\
&X|\theta \sim \text{Bin}(n, \theta)
\end{align*}
\]

Consider tossing the coin a further \(m\) times and let \(z\) be the number of heads obtained. Then \((X \| z) \sim \text{Bin}(n, \theta)\) and:

\[
f(z|x) = \int_0^1 f(z|\theta) f(\theta|x) \, d\theta
\]

\[
= \left( \binom{m}{z} \right) \frac{1}{\beta(a,b)} \int_0^1 \theta^{a+z-1} (1-\theta)^{b+m-z-1} \, d\theta
\]

\[
= \left( \binom{m}{z} \right) \frac{\beta(a+z, b+m-z)}{\beta(a,b)}
\]

\[
= \left( \binom{m}{z} \right) \frac{\Gamma(a+z) \Gamma(b+m-z)}{\Gamma(a+b) \Gamma(z)}
\]

\[
\approx \left( \binom{m}{z} \right) \frac{1}{\Gamma(a+b)} \Gamma(z) \tau(z) \quad z = 0, 1, \ldots, m
\]

This is called the **Binomial-Beta distribution** parameters \(a, b, m\) [also called the Beta-Binomial!].

The predictive distribution can be difficult to calculate; work with but predictive summaries are often easier.
\[ E(\tau) = E(E(\tau|X)) \text{ and condition everywhere on } X \]

1. \( E(\tau | X) = E(\frac{E(\tau | X)}{X}) \)
   \[ = E\left(\frac{E(\tau | X)}{X}\right) \quad \text{if } (\tau | X) \sim \theta \]
   \[
   \begin{align*}
   &\text{calculate using } f(\tau | X) \\
   &\text{calculate using } f(\theta | X)
   \end{align*}
   \]

2. \( \text{Var}(\tau | X) = \text{Var}(E(\tau | X)) + E(\text{Var}(\tau | X)) \)
   \[ = \text{Var}(E(\tau | X)) + E(\text{Var}(\tau | X)) \quad \text{if } (\tau | X) \sim \theta \]
   \[
   \begin{align*}
   &\text{use } f(\tau | X) \\
   &\text{use } f(\theta | X)
   \end{align*}
   \]

**Example: Binomial-Beta.**

\[ E(\tau | X) = E(m | X) \quad \text{as } \tau | X \sim \text{Bin}(m, \theta) \]
\[ = m E(\theta | X) \]
\[ = m \frac{\alpha}{a + b} = m \frac{\alpha + x}{\alpha + \beta + n} \quad \text{as } \theta | X \sim \text{Beta}(\alpha + x, \beta + n - x) \]

\[ \text{Var}(\tau | X) = \text{Var}(m | X) + E(\text{Var}(\theta | X)) \]
\[ = m^2 \text{Var}(\theta | X) + m \left[ E(\theta | X) - E(\theta^2 | X) \right] \]
\[ = \frac{m^2 \alpha \beta}{(a + b)^2(a + b + 1)} + m \left[ \frac{\alpha}{a + b} - \frac{\alpha(a + 1)}{(a + b)(a + b + 1)} \right] \]

Where \( a = \alpha + x, \quad b = \beta + n - x \).

**Note:** \( X \) and \( \tau \) are not independent and nor would we expect them to be. If we don't know \( \theta \) then we expect observing \( X = x \) will be informative for \( \tau \).
Consider such a model from a classical perspective: We view $x$ and $z$ as a random sample where each Bernoulli trial was i.i.d. As we see from the prediction of $z$ given $x = x_*\$.

This is slightly misleading: they are only independent conditional on the parameter $\theta$.

2.2 Exchangerability.

One of the key differences between the classical and Bayesian schools is that observations the former treat as i.i.d are treated as exchangeable by the latter.

The concept of exchangeability was introduced by Bruno de Finetti in the 1930s.

Definition. The RVs $x_1, \ldots, x_n$ are judged to be finitely exchangeable if the joint density function satisfies

$$f(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

for all permutations $\pi$ defined on the set $\{1, \ldots, n\}$.

Hence, in essence, exchangeability captures the notion that only the values matter and not the order in which the observations occur: the labels are uninformative.

NB. Finitely exchangeable implies that

$$f(x_1, \ldots, x_m) = f(x_{\pi(1)}, \ldots, x_{\pi(m)})$$

for all $m \leq n$. 