A prior distribution **must** be specified in a Bayesian analysis.

The choice of prior can substantially affect posterior conclusions, especially when the sample size is not large.

We now examine several broad methods of determining prior distributions.
We know that **conjugacy** is a property of a prior along with a likelihood that implies the posterior distribution will have the same *distributional form* as the prior (just with different parameter(s)).

We have seen some examples of conjugate priors:

<table>
<thead>
<tr>
<th>Data/Likelihood</th>
<th>Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>Beta for $p$</td>
</tr>
<tr>
<td>Poisson</td>
<td>Gamma for $\lambda$</td>
</tr>
<tr>
<td>Normal</td>
<td>Normal for $\mu$</td>
</tr>
<tr>
<td>Normal</td>
<td>Inverse gamma for $\sigma^2$</td>
</tr>
</tbody>
</table>
Conjugate Priors

Other examples:

1. Multinomial $\rightarrow$ Dirichlet for $p_1, p_2, \ldots, p_k$
2. Negative Binomial $\rightarrow$ Beta for $p$
3. Uniform$(0, \theta) \rightarrow$ Pareto for upper limit
4. Exponential $\rightarrow$ Gamma for $\beta$
5. Gamma ($\beta$ unknown) $\rightarrow$ Gamma for $\beta$
6. Pareto ($\alpha$ unknown) $\rightarrow$ Gamma for $\alpha$
7. Pareto ($\beta$ unknown) $\rightarrow$ Pareto for $\beta$
Consider the family of distributions known as the **one-parameter exponential family**.

This family consists of any distribution whose p.d.f. (or p.m.f.) can be written as:

\[
f(x|\theta) = e^{t(x)u(\theta)} r(x)s(\theta)
\]

where \( t(x) \) and \( r(x) \) do not depend on the parameter \( \theta \) and \( u(\theta) \) and \( s(\theta) \) do not depend on \( x \).

Note that any such density can be written as

\[
f(x|\theta) = e^{\{t(x)u(\theta) + \ln[r(x)] + \ln[s(\theta)]\}}
\]
If we observe an iid sample $X_1, \ldots, X_n$, the joint density of the data is thus

$$f(x|\theta) = e^{\left\{u(\theta) \sum_{i=1}^{n} t(x_i) + \sum_{i=1}^{n} \ln[r(x_i)] + n \ln[s(\theta)]\right\}}$$

Consider a prior for $\theta$ (with the prior parameters $k$ and $\gamma$) having the form:

$$p(\theta) = c(k, \gamma) e^{\{k u(\theta) \gamma + k \ln[s(\theta)]\}}$$
Then the posterior is

\[ \pi(\theta | x) \propto f(x | \theta) p(\theta) \]

\[ \propto \exp \left\{ u(\theta) \sum t(x_i) + n \ln[s(\theta)] + ku(\theta) \gamma + k \ln[s(\theta)] \right\} \]

\[ = \exp \left\{ u(\theta) \left[ \sum t(x_i) + k\gamma \right] + (n + k) \ln[s(\theta)] \right\} \]

\[ = \exp \left\{ (n + k)u(\theta) \left[ \frac{\sum t(x_i) + k\gamma}{n + k} \right] + (n + k) \ln[s(\theta)] \right\} \]

which is of the same form as the prior, except with \( k = n + k \) and \( \gamma = \frac{\sum t(x_i) + k\gamma}{n + k} \).

\[ \Rightarrow \] If our data are iid from a one-parameter exponential family, then a conjugate prior will exist.
Conjugate priors are mathematically convenient.

Sometimes they are quite flexible, depending on the specific hyperparameters we use.

But they reflect very specific prior knowledge, so we should be wary of using them unless we truly possess that prior knowledge.
These priors intentionally provide very little specific information about the parameter(s).

A classic uninformative prior is the *uniform* prior.

A *proper* uniform prior integrates to a finite quantity.

**Example 1:** For Bernoulli($\theta$) data, a uniform prior on $\theta$ is

$$p(\theta) = 1, \quad 0 \leq \theta \leq 1.$$ 

This makes sense when $\theta$ has **bounded support**.
Uninformative Priors

Example 2: Consider $N(0, \sigma^2)$ data. If it is “reasonable” to assume, that, say $\sigma^2 < 100$, we could use the uniform prior

$$p(\sigma^2) = \frac{1}{100}, \quad 0 \leq \sigma^2 \leq 100$$

(even though $\sigma^2$ is not intrinsically bounded).

An improper uniform prior integrates to $\infty$:

Example 3: $N(\mu, 1)$ data with

$$p(\mu) = 1, \quad -\infty < \mu < \infty.$$  

This is fine as long as the resulting posterior is proper.

But be careful: Sometimes an improper prior will yield an improper posterior.