

Conjugate priors

A conjugate prior has the same functional form in q as the likelihood function which leads to a posterior distribution belonging to the same distribution family as the prior. For example, the Beta($a_1, a_2, 1$) distribution has probability mass function $f(q)$ given by:

$$f(\theta) = \frac{\theta^{\alpha_1-1} \cdot (1-\theta)^{\alpha_2-1}}{\int_0^1 t^{\alpha_1-1} \cdot (1-t)^{\alpha_2-1} \cdot dt}$$

The denominator is a constant for particular values of a_1 and a_2 so we can rewrite the equation as:

$$f(\theta) \propto \theta^{\alpha_1-1} (1-\theta)^{\alpha_2-1} \tag{1}$$

If we had observed s successes in n trials and were attempting to estimate the true probability of success p , the likelihood function $l(s, n, q)$ would be given by the binomial distribution probability mass function written (using q to represent the unknown parameter p):

$$l(s, n, \theta) = \binom{n}{s} \cdot \theta^s \cdot (1-\theta)^{n-s}$$

Since the binomial coefficient $\binom{n}{s}$ is constant for the given data set (i.e. known n, s), we can rewrite the equation as:

$$l(s, n, \theta) \propto \theta^s \cdot (1-\theta)^{n-s} \tag{2}$$

We can see that the Beta distribution and the binomial likelihood function have the same functional form in q , i.e. $q^a \cdot (1-q)^b$, where a and b are constants. Since the posterior distribution is a product of the prior and likelihood function, it too will have the same functional form, i.e. combining Equations 1 and 2 we have:

$$f(\theta|s, n) \propto \theta^{\alpha_1-1+s} \cdot (1-\theta)^{\alpha_2-1+n-s} \tag{3}$$

We know from the form that this is a Beta(a_1+s, a_2+n-s) distribution, so the posterior density is actually:

$$f(\theta|s, n) = \frac{\theta^{\alpha_1-1+s} \cdot (1-\theta)^{\alpha_2-1+n-s}}{\int_0^1 t^{\alpha_1-1+s} \cdot (1-t)^{\alpha_2-1+n-s} \cdot dt}$$

With a bit of practice, one starts to recognize distributions because of their functional form, without having to go through the step of obtaining the normalized equation. Thus if one uses a Beta distribution as a prior for ρ with a binomial likelihood function, the posterior distribution is also a Beta. The value of using conjugate priors is that we can avoid actually doing any of the mathematics and get directly to the posterior distribution by simply updating the parameters of the prior distribution. Conjugate priors are often called *convenience priors* for obvious reasons.

The Beta(1, 1, 1) distribution is the same as a Uniform(0, 1) distribution, so if we want to start with a Uniform(0, 1) prior for ρ , which makes intuitive sense and also mathematical sense from the viewpoint of [MaxEnt](#), our posterior distribution is given by Beta($s+1, n-s+1, 1$). This is a particularly useful result in [modeling binomial processes](#). The Jeffreys prior for a binomial probability is a Beta($1/2, 1/2, 1$), which peaks at zero and one, but holds to one philosophy of an uninformative prior. Some modelers using a Beta(0, 0, 1) prior which is mathematically undefined and therefore meaningless by itself, giving a posterior distribution of Beta($s, n-s$) which has a mean of s/n . In other words it provides an [unbiased](#) estimate for the binomial probability (a property many statisticians prefer), but has a mode of $(s-1)/(n-2)$ which is not intuitive, and doesn't work if $s=0$ or n .

The following table lists other conjugate priors and the associated likelihood functions. [Exponential families of distributions](#), from which one often draws the likelihood function, all have conjugate priors so the technique can be used frequently in practice. Conjugate priors are also often used to provide approximate but very convenient representations to [subjective priors](#).

| Table of likelihood functions and their conjugate distributions | | | | |
|---|--|--------------------------------------|-------------------------------------|--|
| Likelihood functions | Information | Estimated parameter | Prior | Posterior |
| Multinomial | s_1, s_2, \dots, s_k successes in k categories | Probabilities p_1, p_2, \dots, p_k | Dirichlet(a_1, a_2, \dots, a_k) | $\alpha'_k = \alpha_k + s_k$ |
| Binomial | s successes in n trials | Probability ρ | Beta($a_1, a_2, 1$) | $\alpha'_1 = \alpha_1 + s$ $\alpha'_2 = \alpha_2 + n - s$ |
| Exponential | n "times" x_i | mean ⁻¹ = λ | Gamma($0, b, \lambda$) | $\alpha' = \alpha + n$ $\beta' = \frac{\beta}{1 + \beta \sum_i x_i}$ |
| Normal (with known σ) | n data values with mean \bar{x} | Mean m | Normal(m, σ, n) | $\mu'_\mu = \frac{\mu_\mu(\sigma^2/n) + \bar{x}\sigma_\mu^2}{\sigma^2/n + \sigma_\mu^2}$ $\sigma_\mu = \sqrt{\frac{\sigma_\mu^2 \sigma^2}{n\sigma_\mu^2 + \sigma^2}}$ |
| Poisson | x observations in time t | Mean events per unit time λ | Gamma($0, b, \lambda$) | $\alpha' = \alpha + x$ $\beta' = \frac{\beta}{1 + \beta t}$ |