Stat 700 F'01 Handout on Conjugate Priors

The examples of conjugate priors in natural exponential families are not quite fully general. They have the following character. Suppose that data X is of exponential-family form with density $f_X(x,\vartheta)$ or probability mass function $p_X(x,\vartheta)$ of the form $h(x)\ \exp(\vartheta \cdot \tau(x))$. Note in the examples that follow, $\tau(x)$ is not precisely the same as the minimal sufficient statistic, indeed always has an extra component which is *constant*. This special form requires that the usual term $c(\vartheta)$ itself have an exponential form. Since we have previously observed the general formula

$$\frac{1}{c(\vartheta)} \;=\; \int e^{\vartheta\cdot\tau(x)}\, h(x)\, dx \;=\; E_h\Big(e^{\vartheta\cdot\tau(X)}\Big)$$

with X distributed according to the density $h(x) = f_X(x,0)$, we can see that this special situation corresponds to a moment generating function for $\tau(X)$ under $h(x)$ whose logarithm is linear in ϑ. This in turn is related to the property of *infinite divisibility* for the random vector $\tau(X)$, a concept studied in advanced courses in Probability Theory (like Stat 600-601, under the general heading of 'characteristic functions').

Proposition. Under the assumptions above, if the space Θ of allowed ϑ values is such that the family of functions

$$g(\vartheta,\,\rho) \;=\; k(\rho)\,\exp(\vartheta\,\cdot\,\rho) \qquad (*)$$

can be normalized by constants $k(\rho)$ to be densities with respect to the fixed integrating measure $d\vartheta$ on Θ, then the densities (*) form a family of conjugate priors to the exponential family of densities of X parameterized by ϑ.

Proof. *We do all the steps below with densities, as though X were continuous. Everything goes through analogously for discrete X and pmf's. But note that in all cases, ϑ is treated as a continuous r.v.* When ϑ is regarded as a random variable with density (*), the joint density of (X,ϑ) is $k(\rho)\,h(x)\,\exp(\vartheta\cdot(\tau(x)+\rho)$. Then the marginal density of X is given by

$$f_X(x) \;=\; k(\rho)\,h(x)\int \exp(\vartheta\cdot(\tau(x)+\rho)\,q(\vartheta)\,d\vartheta \;=\; k(\rho)\,h(x)\,/\,k(\rho+t(x))$$

where the last equality follows from the fact that (*) with ρ replaced by $\rho+\tau(x)$ must be a density (when integrated against $d\vartheta$). As a consequence, the posterior density is

$$f_{\vartheta|X}(\vartheta|x) \;=\; k(\rho + \tau(x)) \exp(\vartheta \cdot (\tau(x) + \rho)$$

which is again of the form (*). The Bayesian updating rule for parameters based on observing data x is then $\rho \mapsto \rho + \tau(x)$. The proof is complete. $\square$

The commonest examples of this Proposition are the following:

- *Binomial*: Here $X \sim \binom{n}{x} \exp(x \ln p + (n-x)\ln(1-p))$, and $\tau(x) = (x, n-x)$, $\vartheta = (\ln p, \ln(1-p))$, and the conjugate priors (in terms of p, not ϑ) are proportional to $\exp(\rho_1 \ln p + \rho_2 \ln(1-p))$ with support on $(0,1)$ corresponding to $\Theta = \{(\ln p, \ln(1-p)) : \; p \in (0,1)\}$, i.e., are $Beta(\rho_1 + 1, \rho_2 + 1)$.

- *Poisson*: Now $X \sim (x!)^{-1} \exp(x \ln \lambda - \lambda)$, and $\tau(x) = (x, 1)$, $\vartheta = (\ln \lambda, -\lambda)$, $\Theta = (0, \infty)$. The conjugate priors (in terms of λ) are proportional to $\exp(\rho_1 \ln \lambda - \rho_2 \lambda)$, i.e. are $Gamma(\rho_1 + 1, \rho_2)$.

- *Normal*: If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\tau(x) = (x, x^2, 1)$, and $\vartheta = (\mu/\sigma^2, \, -1/(2\sigma^2), \, -\mu^2/(2\sigma^2))$, and the prior (for μ alone, with σ non-random) is proportional to $\exp(\rho_1 \mu/\sigma^2 - (\rho_2 + \rho_3 \mu^2)/(2(\sigma^2))$, which works as a density only if $\rho_2 = \rho_1^2/\rho_3$ with $\rho_3 > 0$, and then is $\mathcal{N}(\rho_1/\rho_3, \sigma^2/\rho_3)$.

Although they are less commonly mentioned, two further examples, **for which you are to find the conjugate prior families as the first problem on the Take-Home Test**, are :

- *Negative Binomial*: discrete $X \sim NegBin(r, p)$ (in terms of unknown $p \in (0,1)$, with r fixed), and

- *Gamma* : continuous $X \sim Gamma(\alpha, \beta)$ (in terms of unknown $\beta > 0$, with α fixed).