Some Conjugate Distributions

(for continuous random variables)

October 2, 2017

Let the data be $D = (x_1, \ldots, x_n)$.

σ^2 fixed, known

Likelihood:

$$p(D|\mu) \propto n(\bar{x}, \frac{\sigma^2}{n})$$

Prior:

$$\pi(\mu) \propto n(\mu_0, \sigma_0^2)$$

Posterior:

$$\pi(\mu|D) \propto p(D|\mu)\pi(\mu) \propto n(\mu_n, \lambda_n),$$

where

$$\mu_n = \frac{\bar{x}n\lambda + \mu_0\lambda_0}{\lambda_n},$$

$$\lambda_n = \lambda_0 + n\lambda,\ \ \lambda_0 = \sigma_0^{-2},\ \ \lambda = n\sigma^{-2}.$$

Posterior Predictive:

$$p(x|D) = \int p(x|\mu)\pi(\mu|D)d\mu = n(x|\mu_n, \sigma_n^2 + \sigma^2)$$

Both $\mu,\ \lambda = \sigma^{-2}$ unknown

Likelihood:

$$p(x_i|\mu, \lambda) \propto n(\mu, \lambda)$$

$$\Rightarrow p(D|\mu, \lambda) \propto \frac{\lambda^{n/2}}{(2\pi)^{n/2}} exp\{-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\}$$

Prior:

$$\pi(\mu, \lambda) \sim NG(\mu_0, k_0, \alpha_0, \beta_0)$$

$$\stackrel{def}{=} n(\mu|\mu_0, (k_0\lambda)^{-1})Ga(\lambda|\alpha_0, rate = \beta_0)$$

Posterior:

$$\pi(\mu, \lambda|D) = NG(\mu, \lambda|\mu_n, k_n, \alpha_n, \beta_n),$$

where

$$\mu_n = \frac{k_0\mu_0 + n\bar{x}}{k_0 + n},\ k_n = k_0 + n,\ \alpha_n = \alpha_0 + n/2,$$

and

$$\beta_n = \beta_0 + 1/2\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{k_0 n(\bar{x} - \mu_0)^2}{2(k_0 + n)}.$$

Posterior Predictive:

$$p(x|D) = t_{2\alpha_n}(x|\mu_n, \frac{\beta_n(k_n + 1)}{\alpha_n k_n}).$$

μ known, $\lambda = \sigma^{-2}$ unknown

Likelihood:

$$p(x_i|\lambda) \sim n(\mu, \lambda)$$

$$\Rightarrow p(D|\lambda) \propto \lambda^{n/2} exp\{-\frac{\lambda}{2}\sum_{i=1}^{n}(x_i - \mu)^2\}$$

Prior:

$$\pi(\lambda) = Ga(\lambda|\alpha, \beta) \propto \lambda^{\alpha-1} exp\{-\lambda\beta\}$$

Posterior:

$$\pi(\lambda|D) = Ga(\lambda|\alpha_n, \beta_n),$$

where

$$\alpha_n = \alpha + n/2,\ \beta_n = \beta + 1/2\sum_{i=1}^{n}(x_i - \mu)^2$$

Posterior Predictive:

$$\pi(x|D) = t_{2\alpha_n}(x|\mu, \sigma^2 = \frac{\beta_n}{\alpha_n}).$$

Multivariate μ unknown, Σ known
Likelihood:

$$p(D|\mu, \Sigma) \propto n(\bar{x}|\mu, \frac{1}{n}\Sigma)$$

Prior:

$$\pi(\mu) \sim n(\mu_0, \Sigma_0)$$

Posterior:

$$\pi(\mu|D, \Sigma) \sim n(\mu_n, \Sigma_n),$$

where

$$\mu_n = \Sigma_n(n\Sigma^{-1}\bar{x} + \Sigma_0^{-1}\mu_0),\ \ \Sigma_n = (\Sigma_0^{-1} + n\Sigma^{-1})^{-1}$$

Posterior Predictive:

$$\pi(x|D) = n(x|\mu_n, \Sigma + \Sigma_n).$$

Standard Distributions
Gamma:

$$Ga(x|shape = \alpha, rate = \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} exp\{-x\beta\}, x, \alpha, \beta > 0$$

$$Ga(x|shape = \alpha, scale = \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} exp\{-x/\beta\}$$

Using rate parameterization, mean=α/β, mode=$(\alpha-1)/\beta$ for $\alpha \geq 1$ and variance=α/β^2.

If $X \sim Ga(shape = \alpha, rate = \beta)$ and $Y = 1/X$ then $Y \sim IG(shape = \alpha, scale = \beta)$, where

$$IG(x|shape = \alpha, scale = \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} exp\{-\beta/x\},\ \ x, \alpha, \beta > 0,$$

and mean=$\beta/(\alpha-1),\ \alpha > 1$, mode=$\beta/(\alpha+1)$, variance=$\beta^2/[(\alpha-1)^2(\alpha-2)],\ \alpha > 2$.

$$IG(shape = \alpha, rate = \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{-(\alpha+1)} exp\{-1/(\beta x)\},\ \ x, \alpha, \beta > 0.$$

Scaled-Inverse-Chi-Squared distribution:

$$\chi^{-2}(x|\nu,\sigma^2) = \frac{1}{\Gamma(\nu/2)}\left(\frac{\nu\sigma^2}{2}\right)^{\nu/2} x^{-\nu/2-1} exp\{-\frac{\nu\sigma^2}{2x}\},\ x>0$$
$$= IG(shape=\nu/2, scale=\nu\sigma^2/2).$$

Generalized t-Distribution:

$$t_\nu(x|\mu,\sigma^2) = c\left[1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right]^{(\nu+1)/2},\ \nu>0, \sigma^2>0,$$

where

$$c = \frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)}\frac{1}{\sqrt{\nu\pi}\sigma},$$

mean=μ, $\nu > 1$, mode=μ and variance=$\nu\sigma^2/(\nu-2)$, $\nu > 2$. If $x \sim t_\nu(\mu,\sigma^2)$ then $(x-\mu)/\sigma \sim t_\nu$.

It can be shown that the t-distribution is like an infinite sum of Gaussians with different precisions:

$$p(x|\mu,\alpha,\beta) = \int n(x|\mu\tau^{-1})Ga(\tau|\alpha, rate=\beta)d\tau$$
$$= t_{2\alpha}(x|\mu,\beta/\alpha).$$

Multivariate t-distribution in d-dimensions:

$$t_\nu(x|\mu,\Sigma) = \frac{\Gamma(\nu/2+d/2)}{\Gamma(\nu/2)}\frac{|\Sigma|^{-1/2}}{\nu^{d/2}\pi^{d/2}}\left[1+\frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]^{-(\nu+d)/2},$$

where mean=μ, $\nu > 1$, mode=μ and covariance$(x) = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$.

If $Y \sim t(\mu,\Sigma,\nu)$ then marginals are $Y_i \sim t(\mu_i,\Sigma_{ii},\nu)$ and if $Y = (Y_1, Y_2)$ then conditioned distribution is

$$Y_1|y_2 \sim t(\mu_{1|2},\Sigma_{1|2},\nu+d_1),$$

where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(y_2 - \mu_2),$$
$$\Sigma_{1|2} = h_{1|2}\left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T\right),$$

and

$$h_{1|2} = \frac{1}{\nu + d_2}\left[\nu + (y_2 - \mu_2)^T\Sigma_{22}^{-1}(y_2 - \mu_2)\right].$$

We can sample $y \sim t(\mu, \Sigma, \nu)$ by sampling $x \sim t(0, 1, \nu)$ and transforming by $y = \mu + R^T x$ where $R = chol(\Sigma)$ such that $R^T R = \Sigma$.

References:

C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
Bernard and Smith, *Bayesian Theory*, John Wiley & Sons, 1994.
G. Koop, *Bayesian Econometrics*, John Wiley & Sons, 2003.
A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari and D.B. Rubin, *Bayesian Data Analysis, 3rd Edition*, Chapman & Hall/CRC Press, 2013.