Useful Results for Variational Inference. **Matthew J. Beal Summer 2000** (update history 06/12/00, 01/04/02 dd/mm/yy) Nomenclature and layout taken partly from Bayesian Data Analysis, Gelman et al.

Distribution	Notation	Parameters	Density function	Moments, entropy, KL-divergence, etc.	Notes
Uniform	$\theta \sim U(a,b)$ $p(\theta) = U(\theta\|a,b)$	boundaries a,b with $b > a$	$p(\theta) = \frac{1}{b-a}, \theta \in [a,b]$	$H_\theta = \ln(b-a)$ $\langle\theta\rangle = \frac{a+b}{2}, \langle\theta^2\rangle - \langle\theta\rangle^2 = \frac{(b-a)^2}{12}$	
Exponential					
Laplace	$\theta \sim Laplace(\mu,\lambda)$ $p(\theta) = Laplace(\theta\|\mu,\lambda)$	μ mean λ decay scale	$p(\theta) = \frac{1}{2\lambda} e^{-\frac{\|\theta-\mu\|}{\lambda}}$ $\lambda > 0$	$H_\theta = 1 + \ln(2\lambda)$	
Multivariate normal	$\boldsymbol\theta \sim N(\boldsymbol\mu,\Sigma)$ $p(\boldsymbol\theta) = N(\boldsymbol\theta\|\boldsymbol\mu,\Sigma)$	$\boldsymbol\mu$ mean vector Σ covariance	$p(\boldsymbol\theta) = (2\pi)^{-d/2}\|\Sigma\|^{-1/2} e^{-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}(\boldsymbol\theta-\boldsymbol\mu)(\boldsymbol\theta-\boldsymbol\mu)^\top\right]}$	$H_{\boldsymbol\theta} = \frac{d}{2}(\ln 2\pi e) + \frac{1}{2}\ln\|\Sigma\|$ $\mathrm{KL}(\tilde{\boldsymbol\mu},\tilde\Sigma\|\|\boldsymbol\mu,\Sigma) = -\frac{1}{2}\Big(\ln\|\tilde\Sigma\Sigma^{-1}\|$ $+\mathrm{tr}\left[I - \left[\tilde\Sigma + (\tilde{\boldsymbol\mu}-\boldsymbol\mu)(\tilde{\boldsymbol\mu}-\boldsymbol\mu)^\top\right]\Sigma^{-1}\right]\ln e\Big)$ $\langle\boldsymbol\theta\rangle = \boldsymbol\mu$ $\langle\boldsymbol\theta\boldsymbol\theta^\top\rangle = \Sigma$ $K_{\boldsymbol\theta} = \frac{\langle\boldsymbol\theta^4\rangle}{\langle\boldsymbol\theta^2\rangle^2} - 3 = 0$	
Generalised normal			$p(\theta) = \frac{2\beta^{\alpha/2}}{\Gamma(\frac{\alpha}{2})}\theta^{\alpha-1}e^{-\beta\theta^2}$ $\theta,\alpha,\beta > 0$	$H_\theta = \ln\frac{\Gamma(\alpha/2)}{2\beta^{1/2}} - \frac{\alpha-1}{2}\psi(\frac{\alpha}{2}) + \frac{\alpha}{2}$ $\langle\theta\rangle = $ addition $\langle\theta^2\rangle - \langle\theta\rangle^2 = $ addition	
Gamma	$\tau \sim Gamma(\alpha,\beta)$ $p(\tau) = Gamma(\tau\|\alpha,\beta)$	shape $\alpha > 0$ inv. scale $\beta > 0$	$p(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau}$	$H_\tau = \ln\Gamma(\alpha) - \ln\beta + (1-\alpha)\psi(\alpha) + \alpha$ $\langle\tau^n\rangle = \frac{\Gamma(\alpha+n)}{\beta^n\Gamma(\alpha)}$ $\langle(\ln\tau)^n\rangle = \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\partial^n}{\partial\alpha^n}\left(\frac{\Gamma(\alpha)}{\beta^\alpha}\right)$ $\langle\tau\rangle = \alpha/\beta$ $\langle\tau^2\rangle - \langle\tau\rangle^2 = \alpha/\beta^2$ $\langle\ln\tau\rangle = \psi(\alpha) - \ln\beta$ $\mathrm{KL}(\tilde\alpha,\tilde\beta\|\|\alpha,\beta) = \tilde\alpha\ln\tilde\beta - \alpha\ln\beta - \ln\frac{\Gamma(\tilde\alpha)}{\Gamma(\alpha)}$ $+(\tilde\alpha-\alpha)(\psi(\tilde\alpha) - \ln\tilde\beta) - \tilde\alpha(1-\frac{\beta}{\tilde\beta})$	conj. Gaussian precision $y = \tau_1 + \tau_2$ where $\tau_i \sim Gamma(\alpha,\beta)$ then...
Inverse gamma					conj. Gaussian variance
Wishart	$W \sim Wishart_\nu(S)$ $p(W) = Wishart_\nu(W\|S)$	deg. of freedom ν precision matrix S	$p(W) = \frac{1}{Z_{\nu S}}\|W\|^{(\nu-k-1)/2}e^{-\frac{1}{2}\mathrm{tr}\left[S^{-1}W\right]}$ $Z_{\nu S} = 2^{\nu k/2}\pi^{k(k-1)/4}\|S\|^{\nu/2}\prod_{i=1}^k\Gamma\left(\frac{\nu+1-i}{2}\right)$	$H_W = \ln Z_{\nu S} - \frac{\nu-k-1}{2}\langle\ln\|W\|\rangle + \frac{1}{2}\nu k$ $\langle W\rangle = \nu S$ $\langle\ln\|W\|\rangle = \sum_{i=1}^k\psi\left(\frac{\nu+1-i}{2}\right) + k\ln 2 + \ln\|S\|$ $\mathrm{KL}(\tilde\nu,\tilde S\|\|\nu,S) = \ln\frac{Z_{\nu S}}{Z_{\tilde\nu\tilde S}} + \frac{\tilde\nu-\nu}{2}\langle\ln\|W\|\rangle_{\tilde Q} + \frac{1}{2}\tilde\nu\,\mathrm{tr}\left[S^{-1}\tilde S - \mathrm{I}\right]$	conj. Gaussian precision
Inverse-Wishart	$W \sim Inv\text{–}Wishart_\nu(S^{-1})$ $p(W) = Inv\text{–}Wishart_\nu(W\|S)$	deg. of freedom ν covariance matrix S	$p(W) = \frac{1}{Z}\|W\|^{-(\nu+k+1)/2}e^{-\frac{1}{2}\mathrm{tr}\left[SW^{-1}\right]}$ $Z = 2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k\Gamma\left(\frac{\nu+1-i}{2}\right)\times\|S\|^{-\nu/2}$	entropy $\langle W\rangle = (\nu-k-1)^{-1}S$	conj. Gaussian covariance If $W^{-1} \sim Wishart_\nu(S)$, then $W \sim Inv\text{–}Wishart_\nu(S^{-1})$
Student-t (1)	$\theta \sim t_\nu(\mu,\sigma^2)$ $p(\theta) = t_\nu(\theta\|\mu,\sigma^2)$	deg. of freedom $\nu > 0$ mean μ; scale $\sigma > 0$	$p(\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}\left(1 + \frac{1}{\nu}\left(\frac{\theta-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}$	$\langle\theta\rangle = \mu$, for $\nu > 1$ $\langle\theta^2\rangle - \langle\theta\rangle^2 = \frac{\nu}{\nu-2}\sigma^2$, for $\nu > 2$	
Student-t (2)	$\theta \sim t(\mu,\alpha,\beta)$ $p(\theta) = t(\theta\|\mu,\alpha,\beta)$	shape $\alpha > 0$; mean μ scale^2 $\beta > 0$	$p(\theta) = \frac{\Gamma(\alpha+1/2)}{\Gamma(\alpha)\sqrt{2\pi\beta}}\left(1 + \frac{(\theta-\mu)^2}{2\beta}\right)^{-(\alpha+1/2)}$	$H_\theta = \left[\psi(\alpha+\frac{1}{2}) - \psi(\alpha)\right](\alpha+\frac{1}{2}) + \ln\sqrt{2\beta}B(\frac{1}{2},\alpha)$ $K_\theta = \frac{3}{\alpha-2}$ (relative to Gaussian) equiv. $\alpha \to \frac{\nu}{2}; \beta \to \frac{\nu}{2}\sigma^2$	
Multivariate Student-t	$\boldsymbol\theta \sim t_\nu(\boldsymbol\mu,\Sigma)$ $p(\boldsymbol\theta) = t_\nu(\boldsymbol\theta\|\boldsymbol\mu,\Sigma)$	deg. of freedom $\nu > 0$ mean $\boldsymbol\mu$; scale^2 matrix Σ	$p(\boldsymbol\theta) = \frac{1}{Z}\left(1 + \frac{1}{\nu}\mathrm{tr}\left[\Sigma^{-1}(\boldsymbol\theta-\boldsymbol\mu)(\boldsymbol\theta-\boldsymbol\mu)^\top\right]\right)^{-(\nu+d)/2}$ $Z = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)(\nu\pi)^{d/2}}\|\Sigma\|^{-1/2}$	$\langle\boldsymbol\theta\rangle = \boldsymbol\mu$, for $\nu > 1$ $\langle\boldsymbol\theta\boldsymbol\theta^\top\rangle - \langle\boldsymbol\theta\rangle\langle\boldsymbol\theta\rangle^\top = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$	$\nu = 1, t_\nu = Cauchy$ $\nu \to \infty, t_\nu \to N(\boldsymbol\mu,\Sigma)$
Beta	$\theta \sim Beta(\alpha,\beta)$ $p(\theta) = Beta(\theta\|\alpha,\beta)$	prior sample sizes $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$		
Dirichlet	$\boldsymbol\pi \sim Dirichlet(\boldsymbol\alpha)$ $p(\boldsymbol\pi) = Dirichlet(\boldsymbol\pi\|\boldsymbol\alpha)$	prior sample sizes $\boldsymbol\alpha = \{\alpha_1,\ldots,\alpha_k\}$ $\alpha_j > 0; \alpha_0 = \sum_{j=1}^k\alpha_j$	$p(\boldsymbol\pi) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\pi_1^{\alpha_1-1}\cdots\pi_k^{\alpha_k-1}$ $\pi_1,\ldots,\pi_k \geq 0; \sum_{j=1}^k\pi_j = 1$	$\langle\boldsymbol\pi\rangle = \boldsymbol\alpha/\alpha_0$ $\langle\boldsymbol\pi\boldsymbol\pi^\top\rangle - \langle\boldsymbol\pi\rangle\langle\boldsymbol\pi\rangle^\top = \frac{\alpha_0\mathrm{diag}(\boldsymbol\alpha) - \boldsymbol\alpha\boldsymbol\alpha^\top}{\alpha_0^2(\alpha_0+1)}$ $\langle\ln\pi_j\rangle = \psi(\alpha_j) - \psi(\alpha_0)$ $\mathrm{KL}(\tilde{\boldsymbol\alpha},\boldsymbol\alpha) = \ln\frac{\Gamma(\tilde\alpha_0)}{\Gamma(\alpha_0)} - \sum_k\left[\ln\frac{\Gamma(\tilde\alpha_k)}{\Gamma(\alpha_k)} - (\tilde\alpha_k - \alpha_k)(\psi(\tilde\alpha_k) - \psi(\tilde\alpha_0))\right]$ $H_{\boldsymbol\pi} = $ addition	conj. to multinomial
Exponential Family	$\boldsymbol\theta \sim ExpFam_{\boldsymbol\phi}(\boldsymbol\theta)$ $p(\boldsymbol\theta) = ExpFam_{\boldsymbol\phi}(\boldsymbol\theta\|\eta,\boldsymbol\nu)$	number η and value $\boldsymbol\nu$ of pseudo-observations	$p(\boldsymbol\theta) = \frac{1}{Z_{\boldsymbol\phi\eta\boldsymbol\nu}}g(\boldsymbol\theta)^\eta e^{\boldsymbol\phi(\boldsymbol\theta)^\top\boldsymbol\nu}$	$H_{\boldsymbol\theta} = \ln Z_{\boldsymbol\phi\eta\boldsymbol\nu} - \eta\langle\ln g(\boldsymbol\theta)\rangle - \boldsymbol\nu^\top\langle\boldsymbol\phi(\boldsymbol\theta)\rangle$	

Gamma function: $\Gamma(x) = \int_0^\infty d\tau\,\tau^{x-1}e^{-\tau}$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $x! = \Gamma(x+1) = x\Gamma(x) = x(x-1)!$, $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$, $\psi(x) = \frac{\partial}{\partial x}\ln\Gamma(x)$, $\psi(x+1) = \frac{1}{x} + \psi(x)$.