Table 1: Discrete

Bernoulli	p (probability)	Beta	α, β	$\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i$
Binomial	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$
Negative binomial with known failure number, r	p (probability)	Beta	α, β	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + rn$
Poisson	λ (rate)	Gamma	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \ \beta + n$
Multinomial	p (probability vector), k(number of categories)	Dirichlet	α	$\boldsymbol{\alpha} + \sum_{i=1}^{n} \mathbf{x}_i$
Hypergeometric with known total population size, N	M (number of target members)	Beta-binomial	$n = N, \alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$
Geometric	p_0 (probability)	Beta	$\alpha,\ \beta$	$\alpha + n,\ \beta + \sum_{i=1}^{n} x_i - n$

Jan 03, 2019

1 Conjugate Distribution

From Bayes theorem

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta')p(\theta')d\theta'} \tag{1}$$

All membets of the exponential family have conjugate priors.

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters
Bernoulli	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + n - \sum_{i=1}^{n} x_i$
Binomial	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$
Negative binomial with known failure number, r	p (probability)	Beta	$\alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + rn$
Poisson	λ (rate)	Gamma	$k,\ \theta$	$k + \sum_{i=1}^{n} x_i,\ \frac{\theta}{n\theta + 1}$
			$\alpha,\ \beta$ [note 3]	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + n$
Categorical	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + (c_1, \ldots, c_k)$, where c_i is the number of observations in category i
Multinomial	$\boldsymbol{p}$ (probability vector), k (number of categories; i.e., size of $\boldsymbol{p}$)	Dirichlet	$\boldsymbol{\alpha}$	$\boldsymbol{\alpha} + \sum_{i=1}^{n} \mathbf{x}_i$
Hypergeometric with known total population size, N	M (number of target members)	Beta-binomial[4]	$n = N, \alpha,\ \beta$	$\alpha + \sum_{i=1}^{n} x_i,\ \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$
Geometric	p_0 (probability)	Beta	$\alpha,\ \beta$	$\alpha + n,\ \beta + \sum_{i=1}^{n} x_i - n$

Figure 1: Descrete

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters
Normal with known variance σ^2	μ (mean)	Normal	$\mu_0,\, \sigma_0^2$	$\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2}+\frac{\sum_{i=1}^n x_i}{\sigma^2}\right),\left(\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}\right)^{-1}$
Normal with known precision τ	μ (mean)	Normal	$\mu_0,\, \tau_0$	$\frac{\tau_0\mu_0+\tau\sum_{i=1}^n x_i}{\tau_0+n\tau},\, \tau_0+n\tau$
Normal with known mean μ	σ^2 (variance)	Inverse gamma	$\alpha,\, \beta$ [note 5]	$\alpha+\frac{n}{2},\, \beta+\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}$
Normal with known mean μ	σ^2 (variance)	Scaled inverse chi-squared	$\nu,\, \sigma_0^2$	$\nu+n,\, \frac{\nu\sigma_0^2+\sum_{i=1}^n (x_i-\mu)^2}{\nu+n}$
Normal with known mean μ	τ (precision)	Gamma	$\alpha,\, \beta$ [note 3]	$\alpha+\frac{n}{2},\, \beta+\frac{\sum_{i=1}^n (x_i-\mu)^2}{2}$
Normal [note 6]	μ and σ^2 Assuming exchangeability	Normal-inverse gamma	$\mu_0,\, \nu,\, \alpha,\, \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\, \nu+n,\, \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ • $\bar{x}$ is the sample mean
Normal	μ and τ Assuming exchangeability	Normal-gamma	$\mu_0,\, \nu,\, \alpha,\, \beta$	$\frac{\nu\mu_0+n\bar{x}}{\nu+n},\, \nu+n,\, \alpha+\frac{n}{2},$ $\beta+\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}$ • $\bar{x}$ is the sample mean

Figure 2: Continuous 1

Multivariate normal with known covariance matrix $\boldsymbol{\Sigma}$	$\boldsymbol{\mu}$ (mean vector)	Multivariate normal	$\boldsymbol{\mu}_0,\ \boldsymbol{\Sigma}_0$	$\left(\boldsymbol{\Sigma}_0^{-1}+n\boldsymbol{\Sigma}^{-1}\right)^{-1}\left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0+n\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}\right),$ $\left(\boldsymbol{\Sigma}_0^{-1}+n\boldsymbol{\Sigma}^{-1}\right)^{-1}$ • $\bar{\mathbf{x}}$ is the sample mean
Multivariate normal with known precision matrix $\boldsymbol{\Lambda}$	$\boldsymbol{\mu}$ (mean vector)	Multivariate normal	$\boldsymbol{\mu}_0,\ \boldsymbol{\Lambda}_0$	$\left(\boldsymbol{\Lambda}_0+n\boldsymbol{\Lambda}\right)^{-1}\left(\boldsymbol{\Lambda}_0\boldsymbol{\mu}_0+n\boldsymbol{\Lambda}\bar{\mathbf{x}}\right),\ \left(\boldsymbol{\Lambda}_0+n\boldsymbol{\Lambda}\right)$ • $\bar{\mathbf{x}}$ is the sample mean
Multivariate normal with known mean $\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$ (covariance matrix)	Inverse-Wishart	$\nu,\ \boldsymbol{\Psi}$	$n+\nu,\ \boldsymbol{\Psi}+\sum_{i=1}^{n}(\mathbf{x_i}-\boldsymbol{\mu})(\mathbf{x_i}-\boldsymbol{\mu})^T$
Multivariate normal with known mean $\boldsymbol{\mu}$	$\boldsymbol{\Lambda}$ (precision matrix)	Wishart	$\nu,\ \mathbf{V}$	$n+\nu,\ \left(\mathbf{V}^{-1}+\sum_{i=1}^{n}(\mathbf{x_i}-\boldsymbol{\mu})(\mathbf{x_i}-\boldsymbol{\mu})^T\right)^{-1}$
Multivariate normal	$\boldsymbol{\mu}$ (mean vector) and $\boldsymbol{\Sigma}$ (covariance matrix)	normal-inverse-Wishart	$\boldsymbol{\mu}_0,\ \kappa_0,\ \nu_0,\ \boldsymbol{\Psi}$	$\frac{\kappa_0\boldsymbol{\mu}_0+n\bar{\mathbf{x}}}{\kappa_0+n},\ \kappa_0+n,\ \nu_0+n,$ $\boldsymbol{\Psi}+\mathbf{C}+\frac{\kappa_0 n}{\kappa_0+n}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^T$ • $\bar{\mathbf{x}}$ is the sample mean • $\mathbf{C}=\sum_{i=1}^{n}(\mathbf{x_i}-\bar{\mathbf{x}})(\mathbf{x_i}-\bar{\mathbf{x}})^T$
Multivariate normal	$\boldsymbol{\mu}$ (mean vector) and $\boldsymbol{\Lambda}$ (precision matrix)	normal-Wishart	$\boldsymbol{\mu}_0,\ \kappa_0,\ \nu_0,\ \mathbf{V}$	$\frac{\kappa_0\boldsymbol{\mu}_0+n\bar{\mathbf{x}}}{\kappa_0+n},\ \kappa_0+n,\ \nu_0+n,$ $\left(\mathbf{V}^{-1}+\mathbf{C}+\frac{\kappa_0 n}{\kappa_0+n}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)^T\right)^{-1}$ • $\bar{\mathbf{x}}$ is the sample mean • $\mathbf{C}=\sum_{i=1}^{n}(\mathbf{x_i}-\bar{\mathbf{x}})(\mathbf{x_i}-\bar{\mathbf{x}})^T$
Uniform	$U(0,\theta)$	Pareto	$x_m,\ k$	$\max\{\,x_1,\ldots,x_n,x_\mathrm{m}\},\ k+n$

Figure 3: Continuous 2

Pareto with known minimum x_m	k (shape)	Gamma	$\alpha,\ \beta$	$\alpha+n,\ \beta+\sum_{i=1}^n \ln\frac{x_i}{x_{\mathrm{m}}}$
Weibull with known shape β	θ (scale)	Inverse gamma[4]	a, b	$a+n,\ b+\sum_{i=1}^n x_i^{\beta}$
Log-normal with known precision τ	μ (mean)	Normal[4]	$\mu_0,\ \tau_0$	$\left(\tau_0\mu_0+\tau\sum_{i=1}^n \ln x_i\right)\Big/(\tau_0+n\tau),\ \tau_0+n\tau$
Log-normal with known mean μ	τ (precision)	Gamma[4]	$\alpha,\ \beta$[note 3]	$\alpha+\frac{n}{2},\ \beta+\frac{\sum_{i=1}^n(\ln x_i-\mu)^2}{2}$
Exponential	λ (rate)	Gamma	$\alpha,\ \beta$[note 3]	$\alpha+n,\ \beta+\sum_{i=1}^n x_i$
Gamma with known shape α	β (rate)	Gamma	$\alpha_0,\ \beta_0$	$\alpha_0+n\alpha,\ \beta_0+\sum_{i=1}^n x_i$
Inverse Gamma with known shape α	β (inverse scale)	Gamma	$\alpha_0,\ \beta_0$	$\alpha_0+n\alpha,\ \beta_0+\sum_{i=1}^n \frac{1}{x_i}$
Gamma with known rate β	α (shape)	$\propto \frac{a^{\alpha-1}\beta^{\alpha c}}{\Gamma(\alpha)^b}$	$a,\ b,\ c$	$a\prod_{i=1}^n x_i,\ b+n,\ c+n$
Gamma [4]	α (shape), β (inverse scale)	$\propto \frac{p^{\alpha-1}e^{-\beta q}}{\Gamma(\alpha)^r\beta^{-\alpha s}}$	$p,\ q,\ r,\ s$	$p\prod_{i=1}^n x_i,\ q+\sum_{i=1}^n x_i,\ r+n,\ s+n$

Figure 4: Continuous 3