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

Jan 03, 2019

1 Conjugate Distribution

From Bayes theorem

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

All membets of the exponential family have conjugate priors.

2 Inverse Distribution

$$G(y) = \Pr(Y \le y) = \Pr\left(X \ge \frac{1}{y}\right) = 1 - \Pr\left(X < \frac{1}{y}\right) = 1 - F\left(\frac{1}{y}\right)$$

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

Figure 1: Descrete

Likelihood	Model parameters	Conjugate prior distribution	Prior hyperparameters	Posterior hyperparameters
Normal with known variance σ^2	μ (mean)	Normal	μ_0,σ_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	μ_0, au_0	$\frac{\tau_0\mu_0+\tau\sum_{i=1}^nx_i}{\tau_0+n\tau},\tau_0+n\tau$
Normal with known mean μ	σ^2 (variance)	Inverse gamma	α, β ^[note 5]	$lpha+rac{n}{2},eta+rac{\sum_{i=1}^{n}\left(x_{i}-\mu ight)^{2}}{2}$
Normal with known mean μ	σ² (variance)	Scaled inverse chi-squared	$ u, \sigma_0^2 $	$ u+n,rac{ u\sigma_0^2+\sum_{i=1}^n(x_i-\mu)^2}{ u+n}$
Normal with known mean μ	au (precision)	Gamma	α, β ^{note 3]}	$lpha+rac{n}{2},eta+rac{\sum_{i=1}^n(x_i-\mu)^2}{2}$
Normal ^[note 6]	μ and σ^2 Assuming exchangeability	Normal-inverse gamma	$\mu_0, u,lpha,eta$	$\begin{split} &\frac{\nu\mu_0+n\bar{x}}{\nu+n},\nu+n,\alpha+\frac{n}{2},\\ &\beta+\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}\\ &\bullet \bar{x} \text{ is the sample mean} \end{split}$
Normal	μ and τ Assuming exchangeability	Normal- gamma	$\mu_0, u,lpha,eta$	$\begin{split} &\frac{\nu\mu_0+n\bar{x}}{\nu+n},\nu+n,\alpha+\frac{n}{2},\\ &\beta+\frac{1}{2}\sum_{i=1}^n(x_i-\bar{x})^2+\frac{n\nu}{\nu+n}\frac{(\bar{x}-\mu_0)^2}{2}\\ &\bullet\bar{x} \text{ is the sample mean} \end{split}$

Figure 2: Continuous 1

Multivariate normal with known covariance matrix <i>E</i>	μ (mean vector)	Multivariate normal	$oldsymbol{\mu}_0, oldsymbol{\Sigma}_0$	$egin{aligned} \left(oldsymbol{\Sigma}_0^{-1} + noldsymbol{\Sigma}^{-1} ight)^{-1} \left(oldsymbol{\Sigma}_0^{-1}oldsymbol{\mu}_0 + noldsymbol{\Sigma}^{-1}ar{\mathbf{x}} ight), \ \left(oldsymbol{\Sigma}_0^{-1} + noldsymbol{\Sigma}^{-1} ight)^{-1} \end{aligned}$ $oldsymbol{ar{\mathbf{x}}}$ is the sample mean
Multivariate normal with known precision matrix /	μ (mean vector)	Multivariate normal	μ_0, Λ_0	$(oldsymbol{\Lambda}_0 + noldsymbol{\Lambda})^{-1} \left(oldsymbol{\Lambda}_0 oldsymbol{\mu}_0 + noldsymbol{\Lambda}ar{\mathbf{x}} ight), \left(oldsymbol{\Lambda}_0 + noldsymbol{\Lambda} ight)$ $ullet$ $ar{\mathbf{x}}$ is the sample mean
Multivariate normal with known mean μ	Σ (covariance matrix)	Inverse-Wishart	ν, Ψ	$n+ u,\ \mathbf{\Psi}+\sum_{i=1}^n(\mathbf{x_i}-oldsymbol{\mu})(\mathbf{x_i}-oldsymbol{\mu})^T$
Multivariate normal with known mean μ	Λ (precision matrix)	Wishart	ν , V	$n + u$, $\left(\mathbf{V}^{-1} + \sum_{i=1}^{n} (\mathbf{x_i} - \boldsymbol{\mu})(\mathbf{x_i} - \boldsymbol{\mu})^T\right)^{-1}$
Multivariate normal	μ (mean vector) and Σ (covariance matrix)	normal- inverse-Wishart	$oldsymbol{\mu}_0, \kappa_0, u_0, oldsymbol{\Psi}$	$\begin{split} &\frac{\kappa_0 \boldsymbol{\mu}_0 + n \overline{\mathbf{x}}}{\kappa_0 + n}, \kappa_0 + n, \nu_0 + n, \\ &\boldsymbol{\Psi} + \mathbf{C} + \frac{\kappa_0 n}{\kappa_0 + n} (\overline{\mathbf{x}} - \boldsymbol{\mu}_0) (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T \\ &\bullet \overline{\mathbf{x}} \text{ is the sample mean} \\ &\bullet \mathbf{C} = \sum_{i=1}^n (\mathbf{x_i} - \overline{\mathbf{x}}) (\mathbf{x_i} - \overline{\mathbf{x}})^T \end{split}$
Multivariate normal	μ (mean vector) and Λ (precision matrix)	normal-Wishart	$oldsymbol{\mu}_0, \kappa_0, u_0, \mathbf{V}$	$\begin{split} &\frac{\kappa_0 \boldsymbol{\mu}_0 + n \overline{\mathbf{x}}}{\kappa_0 + n}, \kappa_0 + n, \nu_0 + n, \\ &\left(\mathbf{V}^{-1} + \mathbf{C} + \frac{\kappa_0 n}{\kappa_0 + n} (\overline{\mathbf{x}} - \boldsymbol{\mu}_0) (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)^T\right)^{-1} \\ & \bullet \overline{\mathbf{x}} \text{ is the sample mean} \\ & \bullet \mathbf{C} = \sum_{i=1}^n (\mathbf{x_i} - \overline{\mathbf{x}}) (\mathbf{x_i} - \overline{\mathbf{x}})^T \end{split}$
Uniform	$U(0, \theta)$	Pareto	x_m,k	$\max\{x_1,\ldots,x_n,x_{\mathrm{m}}\},k+n$

Figure 3: Continuous 2

		101	157	5 70 1 W
Pareto with known minimum x _m	k (shape)	Gamma	α, β	$lpha+n,eta+\sum_{i=1}^n\lnrac{x_i}{x_{ m m}}$
Weibull with known shape β	heta (scale)	Inverse gamma ^[4]	a, b	$a+n,b+\sum_{i=1}^n x_i^\beta$
Log-normal with known precision $ au$	μ (mean)	Normal ^[4]	μ_0, au_0	$\left(au_0\mu_0+ au\sum_{i=1}^n\ln x_i ight)\Bigg/(au_0+n au), au_0+n au$
Log-normal with known mean μ	τ (precision)	Gamma ^[4]	α , β ^{note 3]}	$lpha+rac{n}{2},eta+rac{\sum_{i=1}^n(\ln x_i-\mu)^2}{2}$
Exponential	λ (rate)	Gamma	α , β ^[note 3]	$\alpha+n,\beta+\sum_{i=1}^n x_i$
Gamma with known shape α	eta (rate)	Gamma	α_0,eta_0	$\alpha_0 + n\alpha, \beta_0 + \sum_{i=1}^n x_i$
Inverse Gamma with known shape α	eta (inverse scale)	Gamma	$lpha_0,eta_0$	$lpha_0 + nlpha, eta_0 + \sum_{i=1}^n rac{1}{x_i}$
Gamma with known rate eta	α (shape)	$\propto rac{a^{lpha-1}eta^{lpha c}}{\Gamma(lpha)^b}$	a, b, c	$a\prod_{i=1}^n x_i,b+n,c+n$
Gamma ^[4]	α (shape), β (inverse scale)	$\propto rac{p^{lpha-1}e^{-eta q}}{\Gamma(lpha)^reta^{-lpha s}}$	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