From
Gelman et al.
(1995)

APPENDIX A

Standard probability distributions

A.1 Introduction

Tables A.1 and A.2 present standard notation, probability density functions, parameter descriptions, and moments for standard probability distributions. The rest of this appendix provides additional information including typical areas of application and methods for simulation.

We use the standard notation θ for the random variable (or random vector), except in the case of the Wishart and inverse-Wishart, for which we use W for the random matrix. The parameters are given conventional labels; all probability distributions are implicitly conditional on the parameters. Most of the distributions here are simple univariate distributions. The multivariate normal and related Wishart and multivariate t, and the multinomial and related Dirichlet distributions, are the principal exceptions. Realistic distributions for complicated multivariate models, including hierarchical and mixture models, can usually be constructed from these building blocks.

For simulating random variables from these distributions, we assume that a computer subroutine or command is available that generates pseudorandom samples from the uniform distribution on the unit interval. Some care must be taken to ensure that the pseudorandom samples from the uniform distribution are appropriate for the task at hand. For example, a sequence may appear uniform in one dimension while m-tuples are not randomly scattered in m dimensions. Many statistical software packages are available for simulating random deviates from the distributions presented here.

A.2 Continuous distributions

Uniform

The uniform distribution is used to represent a variable that is known to lie in an interval and equally likely to be found anywhere in the interval. A noninformative distribution is obtained in the limit as $a \to -\infty$, $b \to \infty$. If u is drawn from a standard uniform distribution $U(0,1)$, then $\theta = a + (b-a)u$ is a draw from $U(a,b)$.

Table A.1 Continuous distributions

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Uniform	$\theta \sim \mathrm{U}(a,b)$ $p(\theta)=\mathrm{U}(\theta\|a,b)$	boundaries a,b with $b>a$	$p(\theta)=\frac{1}{b-a},\ \theta\in[a,b]$	$\mathrm{E}(\theta)=\frac{a+b}{2}$, $\mathrm{var}(\theta)=\frac{(b-a)^2}{12}$ no mode
Normal	$\theta\sim\mathrm{N}(\mu,\sigma^2)$ $p(\theta)=\mathrm{N}(\theta\|\mu,\sigma^2)$	location μ scale $\sigma>0$	$p(\theta)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$\mathrm{E}(\theta)=\mu$, $\mathrm{var}(\theta)=\sigma^2$ $\mathrm{mode}(\theta)=\mu$
Multivariate normal	$\theta\sim\mathrm{N}(\mu,\Sigma)$ $p(\theta)=\mathrm{N}(\theta\|\mu,\Sigma)$ (implicit dimension d)	symmetric, pos. def., $d\times d$ cov. matrix Σ	$p(\theta)=(2\pi)^{-d/2}\|\Sigma\|^{-1/2}$ $\times\exp\left(-\frac{1}{2}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right)$	$\mathrm{E}(\theta)=\mu$, $\mathrm{var}(\theta)=\Sigma$ $\mathrm{mode}(\theta)=\mu$
Gamma	$\theta\sim\mathrm{Gamma}(\alpha,\beta)$ $p(\theta)=\mathrm{Gamma}(\theta\|\alpha,\beta)$	shape $\alpha>0$ inverse scale $\beta>0$	$p(\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\ \theta>0$	$\mathrm{E}(\theta)=\frac{\alpha}{\beta}$ $\mathrm{var}(\theta)=\frac{\alpha}{\beta^2}$ $\mathrm{mode}(\theta)=\frac{\alpha-1}{\beta}$, for $\alpha\geq 1$
Inverse-gamma	$\theta\sim\mathrm{Inv\text{-}gamma}(\alpha,\beta)$ $p(\theta)=\mathrm{Inv\text{-}gamma}(\theta\|\alpha,\beta)$	shape $\alpha>0$ scale $\beta>0$	$p(\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta},\ \theta>0$	$\mathrm{E}(\theta)=\frac{\beta}{\alpha-1}$, for $\alpha>1$ $\mathrm{var}(\theta)=\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha>2$ $\mathrm{mode}(\theta)=\frac{\beta}{\alpha+1}$
Chi-square	$\theta\sim\chi^2_\nu$ $p(\theta)=\chi^2_\nu(\theta)$	deg. of freedom $\nu>0$	$p(\theta)=\frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{\nu/2-1}e^{-\theta/2},\ \theta>0$ same as $\mathrm{Gamma}(\alpha=\frac{\nu}{2},\beta=\frac{1}{2})$	$\mathrm{E}(\theta)=\nu$, $\mathrm{var}(\theta)=2\nu$ $\mathrm{mode}(\theta)=\nu-2$, for $\nu\geq 2$
Inverse-chi-square	$\theta\sim\mathrm{Inv\text{-}}\chi^2_\nu$ $p(\theta)=\mathrm{Inv\text{-}}\chi^2_\nu(\theta)$	deg. of freedom $\nu>0$	$p(\theta)=\frac{2^{-\nu/2}}{\Gamma(\nu/2)}\theta^{-(\nu/2+1)}e^{-1/(2\theta)},\ \theta>0$ same as $\mathrm{Inv\text{-}gamma}(\alpha=\frac{\nu}{2},\beta=\frac{1}{2})$	$\mathrm{E}(\theta)=\frac{1}{\nu-2}$, for $\nu>2$ $\mathrm{var}(\theta)=\frac{2}{(\nu-2)^2(\nu-4)}, \nu>4$ $\mathrm{mode}(\theta)=\frac{1}{\nu+2}$
Scaled inverse-chi-square	$\theta\sim\mathrm{Inv\text{-}}\chi^2(\nu,s^2)$ $p(\theta)=\mathrm{Inv\text{-}}\chi^2(\theta\|\nu,s^2)$	deg. of freedom $\nu>0$ scale $s>0$	$p(\theta)=\frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)}s^\nu\theta^{-(\nu/2+1)}e^{-\nu s^2/(2\theta)},\ \theta>0$ same as $\mathrm{Inv\text{-}gamma}(\alpha=\frac{\nu}{2},\beta=\frac{\nu}{2}s^2)$	$\mathrm{E}(\theta)=\frac{\nu}{\nu-2}s^2$ $\mathrm{var}(\theta)=\frac{2\nu^2}{(\nu-2)^2(\nu-4)}s^4$ $\mathrm{mode}(\theta)=\frac{\nu}{\nu+2}s^2$
Exponential	$\theta\sim\mathrm{Expon}(\beta)$ $p(\theta)=\mathrm{Expon}(\theta\|\beta)$	inverse scale $\beta>0$	$p(\theta)=\beta e^{-\beta\theta},\ \theta>0$ same as $\mathrm{Gamma}(\alpha=1,\beta)$	$\mathrm{E}(\theta)=\frac{1}{\beta}$, $\mathrm{var}(\theta)=\frac{1}{\beta^2}$ $\mathrm{mode}(\theta)=0$
Wishart	$W\sim\mathrm{Wishart}_\nu(S)$ $p(W)=\mathrm{Wishart}_\nu(W\|S)$ (implicit dimension $k\times k$)	deg. of freedom ν symmetric, pos. def. $k\times k$ scale matrix S	$p(W)=\left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{-\nu/2}\|W\|^{(\nu-k-1)/2}$ $\times\exp\left(-\frac{1}{2}\mathrm{tr}(S^{-1}W)\right)$, W pos. def.	$\mathrm{E}(W)=\nu S$
Inverse-Wishart	$W\sim\mathrm{Inv\text{-}Wishart}_\nu(S^{-1})$ $p(W)=\mathrm{Inv\text{-}Wishart}_\nu(W\|S^{-1})$ (implicit dimension $k\times k$)	deg. of freedom ν symmetric, pos. def. $k\times k$ scale matrix S	$p(W)=\left(2^{\nu k/2}\pi^{k(k-1)/4}\prod_{i=1}^k\Gamma\left(\frac{\nu+1-i}{2}\right)\right)^{-1}$ $\times\|S\|^{\nu/2}\|W\|^{-(\nu+k+1)/2}$ $\times\exp\left(-\frac{1}{2}\mathrm{tr}(SW^{-1})\right)$, W pos. def.	$\mathrm{E}(W)=(\nu-k-1)^{-1}S$

Table A.1 **Continuous distributions** *continued*

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Student-t	$\theta \sim t_\nu(\mu,\sigma^2)$ $p(\theta) = t_\nu(\theta\|\mu,\sigma^2)$ t_ν is short for $t_\nu(0,1)$	deg. of freedom $\nu > 0$ location μ scale $\sigma > 0$	$p(\theta) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}\sigma}(1+\frac{1}{\nu}(\frac{\theta-\mu}{\sigma})^2)^{-(\nu+1)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\sigma^2$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Multivariate Student-t	$\theta \sim t_\nu(\mu,\Sigma)$ $p(\theta) = t_\nu(\theta\|\mu,\Sigma)$ (implicit dimension d)	deg. of freedom $\nu > 0$ location $\mu = (\mu_1,\ldots,\mu_d)$ symmetric, pos. def. $d \times d$ scale matrix Σ	$p(\theta) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}}\|\Sigma\|^{-1/2}$ $\times(1+\frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu))^{-(\nu+d)/2}$	$\mathrm{E}(\theta) = \mu$, for $\nu > 1$ $\mathrm{var}(\theta) = \frac{\nu}{\nu-2}\Sigma$, for $\nu > 2$ $\mathrm{mode}(\theta) = \mu$
Beta	$\theta \sim \mathrm{Beta}(\alpha,\beta)$ $p(\theta) = \mathrm{Beta}(\theta\|\alpha,\beta)$	'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$ $\theta \in [0,1]$	$\mathrm{E}(\theta) = \frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ $\mathrm{mode}(\theta) = \frac{\alpha-1}{\alpha+\beta-2}$
Dirichlet	$\theta \sim \mathrm{Dirichlet}(\alpha_1,\ldots,\alpha_k)$ $p(\theta) = \mathrm{Dirichlet}(\theta\|\alpha_1,\ldots,\alpha_k)$	'prior sample sizes' $\alpha_j > 0$; $\alpha_0 \equiv \sum_{j=1}^k \alpha_j$	$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)}\theta_1^{\alpha_1-1}\cdots\theta_k^{\alpha_k-1}$ $\theta_1,\ldots,\theta_k \geq 0; \sum_{j=1}^k \theta_j = 1$	$\mathrm{E}(\theta_j) = \frac{\alpha_j}{\alpha_0}$ $\mathrm{var}(\theta_j) = \frac{\alpha_j(\alpha_0-\alpha_j)}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{cov}(\theta_i,\theta_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$ $\mathrm{mode}(\theta_j) = \frac{\alpha_j-1}{\alpha_0-k}$

Table A.2 **Discrete distributions**

Distribution	Notation	Parameters	Density function	Mean, variance, and mode
Poisson	$\theta \sim \mathrm{Poisson}(\lambda)$ $p(\theta) = \mathrm{Poisson}(\theta\|\lambda)$	'rate' $\lambda > 0$	$p(\theta) = \frac{1}{\theta!}\lambda^\theta \exp(-\lambda)$ $\theta = 0,1,2,\ldots$	$\mathrm{E}(\theta) = \lambda$, $\mathrm{var}(\theta) = \lambda$ $\mathrm{mode}(\theta) = \lfloor\lambda\rfloor$
Binomial	$\theta \sim \mathrm{Bin}(n,p)$ $p(\theta) = \mathrm{Bin}(\theta\|n,p)$	'sample size' n (pos. integer) 'probability' $p \in [0,1]$	$p(\theta) = \binom{n}{\theta}p^\theta(1-p)^{n-\theta}$ $\theta = 0,1,2,\ldots,n$	$\mathrm{E}(\theta) = np$ $\mathrm{var}(\theta) = np(1-p)$ $\mathrm{mode}(\theta) = \lfloor(n+1)p\rfloor$
Multinomial	$\theta \sim \mathrm{Multin}(n;p_1,\ldots,p_k)$ $p(\theta) = \mathrm{Multin}(\theta\|n;p_1,\ldots,p_k)$	'sample size' n (pos. integer) 'probabilities' $p_j \in [0,1]$; $\sum_{j=1}^k p_j = 1$	$p(\theta) = \binom{n}{\theta_1\ \theta_2 \cdots \theta_k}p_1^{\theta_1}\cdots p_k^{\theta_k}$ $\theta_j = 0,1,2,\ldots,n; \sum_{j=1}^k \theta_j = n$	$\mathrm{E}(\theta_j) = np_j$ $\mathrm{var}(\theta_j) = np_j(1-p_j)$ $\mathrm{cov}(\theta_i,\theta_j) = -np_ip_j$
Negative binomial	$\theta \sim \text{Neg-bin}(\alpha,\beta)$ $p(\theta) = \text{Neg-bin}(\theta\|\alpha,\beta)$	shape $\alpha > 0$ inverse scale $\beta > 0$	$p(\theta) = \binom{\theta+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{\beta+1}\right)^\theta$ $\theta = 0,1,2,\ldots$	$\mathrm{E}(\theta) = \frac{\alpha}{\beta}$ $\mathrm{var}(\theta) = \frac{\alpha}{\beta^2}(\beta+1)$
Beta-binomial	$\theta \sim \text{Beta-bin}(n,\alpha,\beta)$ $p(\theta) = \text{Beta-bin}(\theta\|n,\alpha,\beta)$	'sample size' n (pos. integer) 'prior sample sizes' $\alpha > 0, \beta > 0$	$p(\theta) = \frac{\Gamma(n+1)}{\Gamma(\theta+1)\Gamma(n-\theta+1)}\frac{\Gamma(a+\theta)\Gamma(n+b-\theta)}{\Gamma(a+b+n)}$ $\times\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \quad \theta = 0,1,2,\ldots,n$	$\mathrm{E}(\theta) = n\frac{\alpha}{\alpha+\beta}$ $\mathrm{var}(\theta) = n\frac{\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Univariate normal

The normal distribution is ubiquitous in statistical work. Sample averages are approximately normally distributed by the central limit theorem. A noninformative or flat distribution is obtained in the limit as the variance $\sigma^2 \to \infty$. The variance is usually restricted to be positive; $\sigma^2 = 0$ corresponds to a point mass at θ. There are no restrictions on θ. The density function is always finite, the integral is finite as long as σ^2 is finite. A subroutine for generating random draws from the standard normal distribution ($\mu = 0, \sigma = 1$) is available in many computer packages. If not, a subroutine to generate standard normal deviates from a stream of uniform deviates can be obtained from a variety of simulation texts; see Section A.4 for some references. If z is a random deviate from the standard normal distribution, then $\theta = \mu + \sigma z$ is a draw from $\mathrm{N}(\mu, \sigma^2)$.

Two properties of the normal distribution that play a large role in model building and Bayesian computation are the addition and mixture properties. The sum of two independent normal random variables is normally distributed. If θ_1 and θ_2 are independent with $\mathrm{N}(\mu_1, \sigma_1^2)$ and $\mathrm{N}(\mu_2, \sigma_2^2)$ distributions, then $\theta_1 + \theta_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. The mixture property states that if $(\theta_1|\theta_2) \sim \mathrm{N}(\theta_2, \sigma_1^2)$ and $\theta_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$, then $\theta_1 \sim \mathrm{N}(\mu_2, \sigma_1^2 + \sigma_2^2)$. This is useful in the analysis of hierarchical normal models.

Lognormal

If θ is a random variable that is restricted to be positive, and $\log\theta \sim \mathrm{N}(\mu, \sigma^2)$, then θ is said to have a *lognormal* distribution. Using the Jacobian of the log transformation, one can directly determine that the density is $p(\theta) = (\sqrt{2\pi}\sigma\theta)^{-1}\exp(-\frac{1}{2\sigma^2}(\log\theta - \mu)^2)$, the mean is $\exp(\mu + \frac{1}{2}\sigma^2)$, the variance is $\exp(2\mu)\exp(\sigma^2)(\exp(\sigma^2) - 1)$, and the mode is $\exp(\mu - \sigma^2)$.

Multivariate normal

The multivariate normal density is always finite; the integral is finite as long as $\det(\Sigma^{-1}) > 0$. A noninformative distribution is obtained in the limit as $\det(\Sigma^{-1}) \to 0$; this limit is not uniquely defined. A random draw from a multivariate normal distribution can be obtained using the Cholesky decomposition of Σ and a vector of univariate normal draws. The Cholesky decomposition of Σ produces a lower-triangular matrix A (the 'Cholesky factor') for which $AA^T = \Sigma$. If $z = (z_1, \ldots, z_d)$ are d independent standard normal random variables, then $\theta = \mu + Az$ is a random draw from the multivariate normal distribution with covariance matrix Σ.

The marginal distribution of any subset of components (for example, θ_i or (θ_i, θ_j)) is also normal. Any linear transformation of θ, such as the projection of θ onto a linear subspace, is also normal, with dimension equal to the rank of the transformation. The conditional distribution of θ, constrained to lie on any linear subspace, is also normal. The addition property holds: if θ_1 and θ_2 are independent with $\mathrm{N}(\mu_1, \Sigma_1)$ and $\mathrm{N}(\mu_2, \Sigma_2)$ distributions, then $\theta_1 + \theta_2 \sim \mathrm{N}(\mu_1 + \mu_2, \Sigma_1 + \Sigma_2)$ as long as θ_1 and θ_2 have the same dimension. We discuss the generalization of the mixture property shortly.

The conditional distribution of any subvector of θ given the remaining elements is once again multivariate normal. If we partition θ into subvectors $\theta = (U, V)$, then $p(U|V)$ is (multivariate) normal:

$$\begin{aligned} \mathrm{E}(U|V) &= \mathrm{E}(U) + \mathrm{cov}(V,U)\mathrm{var}(V)^{-1}(V - \mathrm{E}(V)), \\ \mathrm{var}(U|V) &= \mathrm{var}(U) - \mathrm{cov}(V,U)\mathrm{var}(V)^{-1}\mathrm{cov}(U,V), \end{aligned} \quad \text{(A.1)}$$

where $\mathrm{cov}(V,U)$ is a rectangular matrix (submatrix of Σ) of the appropriate dimensions, and $\mathrm{cov}(U,V) = \mathrm{cov}(V,U)^T$. In particular, if we define the matrix of conditional coefficients,

$$C = I - [\mathrm{diag}(\Sigma^{-1})]^{-1}\Sigma^{-1},$$

then

$$(\theta_i \,|\, \theta_j, \text{all } j \neq i) \sim \mathrm{N}(\mu_i + \sum_{j \neq i} c_{ij}\mu_j, (\Sigma^{-1})_{ii}). \quad \text{(A.2)}$$

Conversely, if we parametrize the distribution of U and V hierarchically:

$$U|V \sim \mathrm{N}(XV, \Sigma_{U|V}), \quad V \sim \mathrm{N}(\mu_V, \Sigma_V),$$

then the joint distribution of θ is the multivariate normal,

$$\theta = \begin{pmatrix} U \\ V \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} X\mu_V \\ \mu_V \end{pmatrix}, \begin{pmatrix} X\Sigma_V X^T + \Sigma_{U|V} & X\Sigma_V \\ \Sigma_V X^T & \Sigma_V \end{pmatrix}\right).$$

This generalizes the mixture property of univariate normals.

The 'weighted sum of squares,' $SS = (\theta - \mu)^T\Sigma^{-1}(\theta - \mu)$, has a χ^2_k distribution. For any matrix A for which $AA^T = \Sigma$, the conditional distribution of $A^{-1}(\theta - \mu)$, given SS, is uniform on the $(d-1)$-dimensional unit sphere.

Gamma

The gamma distribution is the conjugate prior distribution for the inverse of the normal variance and for the mean parameter of the Poisson distribution. The gamma integral is finite if $\alpha > 0$; the density function is finite if $\alpha \geq 1$. A noninformative distribution is obtained in the limit as $\alpha \to 0$, $\beta \to 0$. Many computer packages generate gamma random variables directly; otherwise, it is possible to obtain draws from a gamma random variable using draws from a uniform as input. The most effective method depends on the parameter α; see the references for details.

There is an addition property for independent gamma random variables with the same inverse scale parameter. If θ_1 and θ_2 are independent with $\mathrm{Gamma}(\alpha_1, \beta)$ and $\mathrm{Gamma}(\alpha_2, \beta)$ distributions, then $\theta_1 + \theta_2 \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \beta)$. The logarithm of a gamma random variable is approximately normal; raising a gamma random variable to the one-third power provides an even better normal approximation.

Inverse-gamma

If θ^{-1} has a gamma distribution with parameters α, β, then θ has the inverse-gamma distribution. The density is finite always; its integral is finite if $\alpha > 0$. The inverse-gamma is the conjugate prior distribution for the normal variance. A noninformative distribution is obtained as $\alpha, \beta \to 0$.

Chi-square

The χ^2 distribution is a special case of the gamma distribution, with $\alpha = \nu/2$ and $\beta = \frac{1}{2}$. The addition property holds since the inverse scale parameter is fixed: if θ_1 and θ_2 are independent with $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$ distributions, then $\theta_1 + \theta_2 \sim \chi^2_{\nu_1+\nu_2}$.

Inverse chi-square

The inverse-χ^2 is a special case of the inverse-gamma distribution, with $\alpha = \nu/2$ and $\beta = \frac{1}{2}$. We also define the *scaled* inverse chi-square distribution, which is useful for variance parameters in normal models. To obtain a simulation draw θ from the Inv-$\chi^2(\nu, s^2)$ distribution, first draw X from the χ^2_ν distribution and then let $\theta = \nu s^2/X$.

Exponential

The exponential distribution is the distribution of waiting times for the next event in a Poisson process and is a special case of the gamma distribution with $\alpha = 1$. Simulation of draws from the exponential distribution is straightforward. If U is a draw from the uniform distribution on $[0,1]$, then $-\log(U)/\beta$ is a draw from the exponential distribution with parameter β.

Weibull

If θ is a random variable that is restricted to be positive, and $(\theta/\beta)^\alpha$ has an Expon(1) distribution, then θ is said to have a *Weibull* distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$. The Weibull is often used to model failure times in reliability analysis. Using the Jacobian of the log transformation, one can directly determine that the density is $p(\theta) = \frac{\alpha}{\beta^\alpha}\theta^{\alpha-1}\exp(-(\theta/\beta)^\alpha)$, the mean is $\beta\Gamma(1+\frac{1}{\alpha})$, the variance is $\beta^2[\Gamma(1+\frac{2}{\alpha}) - (\Gamma(1+\frac{1}{\alpha}))^2]$, and the mode is $\beta(1-\frac{1}{\alpha})^{1/\alpha}$.

Wishart

The Wishart is the conjugate prior distribution for the inverse covariance matrix in a multivariate normal distribution. It is a multivariate generalization of the gamma distribution. The integral is finite if the degrees of freedom parameter, ν, is greater than or equal to the dimension, k. The density is finite if $\nu \geq k+1$. A noninformative distribution is obtained as $\nu \to 0$. The sample covariance matrix for iid multivariate normal data has a Wishart distribution. In fact, multivariate normal simulations can be used to simulate a draw from the Wishart distribution, as follows. Simulate $\alpha_1, \ldots, \alpha_\nu$, ν independent samples from a k-dimensional multivariate $\mathrm{N}(0, S)$ distribution, then let $\theta = \sum_{i=1}^{\nu} \alpha_i \alpha_i^T$. This only works when the distribution is proper; that is, $\nu \geq k$.

Inverse-Wishart

If $W^{-1} \sim \mathrm{Wishart}_\nu(S)$ then W has the inverse-Wishart distribution. The inverse-Wishart is the conjugate prior distribution for the multivariate normal covariance matrix. The inverse-Wishart density is always finite, and the integral is always finite. A degenerate form occurs when $\nu < k$.

Student-t

The t is the marginal posterior distribution for the normal mean with unknown variance and conjugate prior distribution and can be interpreted as a mixture of normals with common mean and variances that follow an inverse-gamma distribution. The t is also the ratio of a normal random variable and the square root of an independent gamma random variable. To simulate t, simulate z from a standard normal and x from a χ^2_ν, then let $\theta = \mu + \sigma z\sqrt{\nu}/\sqrt{x}$. The t density is always finite; the integral is finite if $\nu > 0$ and σ is finite. In the limit $\nu \to \infty$, the t distribution approaches $\mathrm{N}(\mu, \sigma^2)$. The case of $\nu = 1$ is called the *Cauchy distribution*. The t distribution can be used in place of a normal distribution in a robust analysis.

Beta

The beta is the conjugate prior distribution for the binomial probability. The density is finite if $\alpha, \beta \geq 1$, and the integral is finite if $\alpha, \beta > 0$. The choice $\alpha = \beta = 1$ gives the standard uniform distribution; $\alpha = \beta = 0.5$ and $\alpha = \beta = 0$ are also sometimes used as noninformative densities. To simulate θ from the beta distribution, first simulate x_α and x_β from $\chi^2_{2\alpha}$ and $\chi^2_{2\beta}$ distributions, respectively, then let $\theta = \frac{x_\alpha}{x_\alpha + x_\beta}$.

It is sometimes useful to estimate quickly the parameters of the beta distribution using the method of moments:

$$\alpha + \beta = \frac{\mathrm{E}(\theta)(1-\mathrm{E}(\theta))}{\mathrm{var}(\theta)} - 1$$

$$\alpha = (\alpha+\beta)\mathrm{E}(\theta), \qquad \beta = (\alpha+\beta)(1-\mathrm{E}(\theta)). \tag{A.3}$$

The beta distribution is also of interest because the kth order statistic from a sample of n iid $\mathrm{U}(0,1)$ variates has the $\mathrm{Beta}(k, n-k+1)$ distribution.

Dirichlet

The Dirichlet is the conjugate prior distribution for the parameters of the multinomial distribution. The Dirichlet is a multivariate generalization of the beta distribution. As with the beta, the integral is finite if all of the α's are positive, and the density is finite if all are greater than or equal to one. A noninformative prior is obtained as $\alpha_j \to 0$ for all j.

The marginal distribution of a single θ_j is $\text{Beta}(\alpha_j, \alpha_0 - \alpha_j)$. The marginal distribution of a subvector of θ is Dirichlet; for example $(\theta_i, \theta_j, 1-\theta_i-\theta_j) \sim \text{Dirichlet}(\alpha_i, \alpha_j, \alpha_0 - \alpha_i - \alpha_j)$. The conditional distribution of a subvector given the remaining elements is Dirichlet under the condition $\sum_{j=1}^k \theta_j = 1$.

There are two standard approaches to sampling from a Dirichlet distribution. The fastest method generalizes the method used to sample from the beta distribution: draw $x_1, \dots, x_k$ from independent gamma distributions with common scale and shape parameters $\alpha_1, \dots, \alpha_k$, and for each j, let $\theta_j = x_j / \sum_{i=1}^k x_i$. A less efficient algorithm relies on the univariate marginal and conditional distributions being beta and proceeds as follows. Simulate θ_1 from a $\text{Beta}(\alpha_1, \sum_{i=2}^k \alpha_i)$ distribution. Then simulate $\theta_2, \dots, \theta_{k-1}$ in order, as follows. For $j = 2, \dots, k-1$, simulate ϕ_j from a $\text{Beta}(\alpha_j, \sum_{i=j+1}^k \alpha_i)$ distribution, and let $\theta_j = (1 - \sum_{i=1}^{j-1} \theta_i)\phi_j$. Finally, set $\theta_k = 1 - \sum_{i=1}^{k-1} \theta_i$.

A.3 Discrete distributions

Poisson

The Poisson distribution is commonly used to represent count data, such as the number of arrivals in a fixed time period. The Poisson distribution has an addition property: if θ_1 and θ_2 are independent with $\text{Poisson}(\lambda_1)$ and $\text{Poisson}(\lambda_2)$ distributions, then $\theta_1 + \theta_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$. Simulation for the Poisson distribution (and most discrete distributions) can be cumbersome. Table lookup can be used to invert the cumulative distribution function. Simulation texts describe other approaches.

Binomial

The binomial distribution is commonly used to represent the number of 'successes' in a sequence of n iid Bernoulli trials, with probability of success p in each trial. A binomial random variable with large n is approximately normal. If θ_1 and θ_2 are independent with $\text{Bin}(n_1, p)$ and $\text{Bin}(n_2, p)$ distributions, then $\theta_1 + \theta_2 \sim \text{Bin}(n_1 + n_2, p)$. For small n, a binomial random variable can be simulated by obtaining n independent standard uniforms and setting θ equal to the number of uniform deviates less than or equal to p. For larger n, more efficient algorithms are often available in computer packages. When $n = 1$, the binomial is called the *Bernoulli* distribution.

Multinomial

The multinomial distribution is a multivariate generalization of the binomial distribution. The marginal distribution of a single θ_i is binomial. The conditional distribution of a subvector of θ is multinomial with 'sample size' parameter reduced by the fixed components of θ and 'probability' parameters rescaled to have sum equal to one. We can simulate a multivariate draw using a sequence of binomial draws. Draw θ_1 from a $\text{Bin}(n, p_1)$ distribution. Then draw $\theta_2, \dots, \theta_{k-1}$ in order, as follows. For $j = 2, \dots, k-1$, draw θ_j from a $\text{Bin}(n - \sum_{i=1}^{j-1} \theta_i, p_j / \sum_{i=j}^k p_i)$ distribution. Finally, set $\theta_k = n - \sum_{i=1}^{k-1} \theta_i$. If at any time in the simulation the binomial sample size parameter equals zero, use the convention that a $\text{Bin}(0, p)$ variable is identically zero.

Negative binomial

The negative binomial distribution is the marginal distribution for a Poisson random variable when the rate parameter has a $\text{Gamma}(\alpha, \beta)$ prior distribution. The negative binomial can also be used as a robust alternative to the Poisson distribution, because it has the same sample space, but has an additional parameter. To simulate a negative binomial random variable, draw $\lambda \sim \text{Gamma}(\alpha, \beta)$ and then draw $\theta \sim \text{Poisson}(\lambda)$. In the limit $\alpha \to \infty$, and $\alpha/\beta \to$constant, the distribution approaches a Poisson with parameter α/β. Under the alternative parametrization, $p = \frac{\beta}{\beta+1}$, θ can be interpreted as the number of Bernoulli failures obtained before the α successes, where the probability of success is p.

Beta-binomial

The beta-binomial arises as the marginal distribution of a binomial random variable when the probability of success has a $\text{Beta}(\alpha, \beta)$ prior distribution. It can also be used as a robust alternative to the binomial distribution. The mixture definition gives an algorithm for simulating from the beta-binomial: draw $\phi \sim \text{Beta}(\alpha, \beta)$ and then draw $\theta \sim \text{Bin}(n, \phi)$.

A.4 Bibliographic note

Many software packages contain subroutines to simulate draws from these distributions. Texts on simulation typically include information about many of these distributions; for example, Ripley (1987) discusses simulation of all of these in detail, except for the Dirichlet and multinomial. Johnson and Kotz (1972) give more detail, such as the characteristic functions, for the distributions. Fortran and C programs for uniform, normal, gamma, Poisson, and binomial distributions are available in Press et al. (1986).