EC 607: Bayesian Econometrics

Review of Important Probability Density Functions

Prof. Jeremy M. Piger

Department of Economics

University of Oregon

Last Revised: March 20, 2018

The Bayesian approach to econometrics makes extensive use of a number of probability density functions (pdf) and probability mass functions (pmf). Here we will discuss the most common pdfs and pmfs used in Bayesian econometrics.

1 Location, Scale and Shape Parameters

When working with a parametric family of pdfs, we classify parameters into three types:

- **Location Parameter:** A parameter that shifts the location of a pdf. That is, it moves the pdf along the x-axis, without affecting its shape or dispersion. Formally, suppose we have a random variable x with a pdf that depends on parameters μ and θ given by $p(x; \mu, \theta)$. Then μ is a location parameter if:

$$p(x; \mu, \theta) = p(x - \mu; 0, \theta)$$

- **Scale Parameter:** A parameter that changes how dispersed a pdf is, without changing its shape. Suppose we have a random variable x with a pdf that depends on parameters s and θ given by $p(x; s, \theta)$. Then s is a scale parameter if:

$$p(x; s, \theta) = \frac{1}{s} p(x/s; 1, \theta)$$

- **Shape Parameter:** A parameter of a pdf that is neither a location parameter nor a scale parameter (nor a function of either or both of these only). Such a parameter must affect the shape of a pdf rather than simply shifting it or stretching/shrinking it.

2 Univariate Probability Density Functions

2.1 Uniform Distribution

A continuous scalar random variable x has a **uniform distribution** with parameters a and b if the pdf for x is given by:

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise.} \end{cases},$$

where $a < b$. The typical notation is: $x \sim U(a,b)$.

Summary statistics for uniform distributed random variables include:

Mean:	$\frac{a+b}{2}$
Median:	$\frac{a+b}{2}$
Variance:	$\frac{(b-a)^2}{12}$

The uniform distribution is a common choice as prior pdf for a variety of parameters in econometric models.

In Matlab, the function unifrnd(a,b,m,n) generates an $m \times n$ matrix of independent $U(a,b)$ random variables.

2.2 Normal Distribution

A continuous scalar random variable x has a **normal distribution** with location parameter μ and scale parameter σ if the pdf for x is given by:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

where $\sigma > 0$. The typical notation is: $x \sim N(\mu, \sigma^2)$

Summary statistics for normally distributed random variables include:

Mean:	μ
Median:	μ
Variance:	σ^2

The normal distribution is a common choice as prior pdf for conditional mean parameters in econometric models.

In Matlab, the function normrnd(μ,σ,m,n) generates an $m \times n$ matrix of independent $N(\mu, \sigma^2)$ random variables.

2.3 t-Distribution

A continuous scalar random variable x has a **t-distribution** with location parameter μ, scale parameter σ, and shape parameter v if the pdf for x is given by:

$$p(x) = \frac{v^{v/2}\Gamma\left(\frac{v+1}{2}\right)}{\sigma\pi^{1/2}\Gamma\left(\frac{v}{2}\right)}\left[v + \left(\frac{x-\mu}{\sigma}\right)^2\right]^{-(v+1)/2} \quad \text{for} \ \ v, \sigma > 0.$$

where $\Gamma(\cdot)$ is the Gamma function. The Gamma function is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers. Thus, if α is a positive integer:

$$\Gamma(\alpha) = (\alpha - 1)!$$

This is a useful result that we will use later.

The typical notation is: $x \sim t(\mu, \sigma, v)$. The parameter v is often called the "degrees of freedom parameter."

Summary statistics for t-distributed random variables include:

Mean:	μ for $v > 1$.
Median:	μ
Variance:	$\frac{v}{v-2}\sigma^2$ for $v > 2$.

The t-distribution is sometimes used as prior pdf for conditional mean parameters in econometric models.

In Matlab, the function $trnd(v, m, n)$ generates an $m \times n$ matrix of independent $t(0, 1, v)$ random variables. Such realizations are said to be distributed as "Student t". These can be turned into $t(\mu, \sigma^2, v)$ random variables by multiplying each random variable realization by σ and then adding μ.

2.4 Gamma Distribution

A continuous random variable x has a **Gamma distribution** with shape parameter α and scale parameter β if the pdf for x is given by:

$$p(x) = \frac{1}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}x^{\alpha-1}\exp\left[-\frac{x}{\beta}\right] \quad \text{for} \quad x > 0 \quad \text{and} \quad \alpha, \beta > 0,$$

The typical notation is: $x \sim \text{Gamma}(\alpha, \beta)$ or $x \sim G(\alpha, \beta)$.

Summary statistics for Gamma distributed random variables include:

Mean:	$\alpha\beta$
Median:	no closed form expression
Variance:	$\alpha\beta^2$

The Gamma distribution is a common choice as prior pdf for the inverse of conditional variance parameters (known as precision parameters), as well as other parameters constrained to be positive.

In Matlab, the function gamrnd(α,β,m,n) generates an $m \times n$ matrix of independent Gamma(α, β) random variables.

Note that $\frac{1}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}$ doesn't depend on x. Since the pdf must integrate to unity over the possible space for x, this implies:

$$\beta^\alpha \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp\left[-\frac{x}{\beta}\right] dx$$

This is a useful result that we will use later.

Another equivalent formulation of the Gamma distribution that is common in the Bayesian econometrics literature is as follows:

$$p(x) = \frac{1}{\left(\frac{2\mu}{v}\right)^{v/2}} \frac{1}{\Gamma\left(\frac{v}{2}\right)} x^{\frac{v-2}{2}} \exp\left[-\frac{xv}{2\mu}\right] \quad \text{for} \quad x > 0 \quad \text{and} \quad \mu, v > 0,$$

The two formulations are related by $v = 2\alpha$ and $\mu = \alpha\beta$. So, for this formulation, the summary statistics are given by:

Mean:	μ
Median:	no closed form expression
Variance:	$2\mu^2/v$

There are two other well-known distributions that are special cases of the Gamma distribution. The **Chi-square distribution** is a Gamma distribution with $v = \mu$. An **Exponential distribution** is a Gamma distribution with $v = 2$.

2.5 Inverse Gamma Distribution

If a continuous random variable $x \sim \text{Gamma}(\alpha, \beta)$, then $z = 1/x$ will have an **Inverse Gamma distribution** with shape parameter α and scale parameter $1/\beta$. Its pdf is given by:

$$p(z) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}} \text{ for } x > 0 \text{ and } \alpha, \beta > 0,$$

where $\Gamma(\cdot)$ is the Gamma function.

The typical notation is: $x \sim \text{IG}(\alpha, 1/\beta)$. The inverse gamma distribution can be derived via a straightforward change of variables transformation of the Gamma distribution pdf.

The Inverse Gamma distribution is a common choice as prior pdf for variance parameters. It's use can be avoided if one works with "precision" parameters rather than variance parameters. A precision parameter is the inverse of a variance parameter.

Summary statistics for Inverse Gamma distributed random variables include:

Mean:	$\frac{\beta}{(\alpha-1)}$ for $\alpha > 1$
Median:	no closed form expression
Variance:	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$

A random variable that is distributed $IG(\alpha, 1/\beta)$ can be generated by generating a random variable that is distributed $G(\alpha, \beta)$ then taking the reciprocal of this generated random variable.

2.6 Binomial and Bernoulli Distributions

A discrete random variable x has a **Binomial distribution** with parameters T and θ if the pmf for x is given by:

$$\Pr(x) = \begin{cases} \binom{T}{x} \theta^x (1-\theta)^{T-x} & \text{if } x = 0, 1, 2, \cdots, T \\ 0 & \text{if } x < 0 \end{cases},$$

where $0 < \theta < 1$ and T is a positive integer.

The typical notation is: $x \sim B(T, \theta)$.

Summary statistics for Binomial distributed random variables include:

Mean:	$T\theta$
Median:	$\lfloor T\theta \rfloor$ or $\lceil T\theta \rceil$
Variance:	$T\theta(1-\theta)$

The Binomial distribution is used as a likelihood function where an experiment results in either "success" or "failure", and is repeated independently T times, with probability of success for each experiment given by θ. The distribution of the random variable x, which counts the number of successes, is $B(T, \theta)$.

The **Bernoulli** distribution is a special case of the Binomial distribution with $T = 1$. For this reason, the T experiments referenced in the Binomial distribution are sometimes referred to as "Bernoulli trials."

In Matlab, the function binornd(T, θ, m, n) generates an $m \times n$ matrix of independent $B(T, \theta)$ random variables.

2.7 Poisson Distribution

A discrete random variable x has a **Poisson distribution** with shape parameter λ if the pmf for x is given by:

$$\Pr(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & \text{if } x = 0, 1, 2, \cdots, \\ 0 & \text{otherwise} \end{cases},$$

where $\lambda > 0$.

The typical notation is: $x \sim P_o(\lambda)$.

Summary statistics for a Poisson distributed random variable includes:

Mean:	λ
Median:	No exact solution
Variance:	λ

The Poisson distribution is a popular choice of likelihood function for "count data," where a random variable denotes the number of times an event occurs. Note that the binomial distribution is also a model for counts, but it assumes an upper limit (T in the notation above) on the value of the count. The Poisson distribution does not make such an assumption.

In Matlab, the function poissrnd(λ, m, n) generates an $m \times n$ matrix of independent $P_o(\lambda)$ random variables.

2.8 Negative Binomial Distribution

Note that for a Poisson distributed random variable, its mean equals its variance. However, in actual count data the variance is often far larger than the mean, a situation referred to as "overdispersion."

When count data is overdispersed, a popular alternative to the Poisson is the **Negative Binomial Distribution**.

A discrete random variable x has a Negative Binomial distribution with parameters r and θ if the pmf for x is given by:

$$\Pr(x) = \begin{cases} \binom{x+r-1}{x} \theta^x (1-\theta)^r & \text{if } x = 0, 1, 2, \cdots \\ 0 & \text{if } x < 0 \end{cases},$$

The negative binomial distribution can be interpreted as follows. Suppose we do a series of Bernoulli trials, with success probability θ, stopping when r failures have occurred. The number of successes, x, will then have a negative binomial distribution.

The typical notation is: $x \sim \mathrm{NB}(r, \theta)$.

Summary statistics for Negative Binomial random variables include:

Mean:	$\frac{\theta r}{1-\theta}$
Median:	no closed form expression
Variance:	$\frac{\theta r}{(1-\theta)^2}$

The negative binomial distribution can be extended to non-integer r. In this case, the formula for the pmf is:

$$\Pr(x) = \begin{cases} \frac{\Gamma(x+r)}{\Gamma(x+1)\Gamma(r)} \theta^x (1-\theta)^r & \text{if } x = 0, 1, 2, \cdots \\ 0 & \text{if } x < 0 \end{cases},$$

In this case, the negative binomial can still be thought of as a discrete counting distribution, but it does not have the interpretation given above in terms of Bernoulli trials (since r is not an integer).

Note that the Negative Binomial distribution can be derived as the pmf of a random variable that arises from a Poisson distribution whose parameter λ is itself a random variable, and is distributed according to a Gamma Distribution. This is an example of a "mixing" distribution. The extra randomness from λ gives the negative binomial an additional source of variance over the Poisson.

3 Multivariate Distributions

3.1 Dirichlet and Beta Distribution

Consider a vector of continuous random variables $X = (x_1, x_2, \cdots, x_N)'$, where $\sum_{i=1}^{N} x_i = 1$. X has a **Dirichlet distribution** with vector of shape parameters $\alpha = (\alpha_1, \alpha_2, \cdot, \cdot\cdot, \alpha_N)'$ if the joint pdf for X is given by:

$$p(X) = \left[\frac{\Gamma(a)}{\prod_{i=1}^{N} \Gamma(\alpha_i)}\right] \prod_{i=1}^{N} x_i^{\alpha_i - 1} \quad \text{for} \quad 0 < x_i < 1 \quad \text{and} \quad \alpha_i > 0,$$

where $a = \sum_{i=1}^{N} \alpha_i$. Note that since $x_N = 1 - \sum_{i=1}^{N-1} x_i$, the Dirichlet is a $N-1$-variate pdf, rather than an N-variate.

The typical notation is: $X \sim D(\alpha)$.

Summary statistics for a Dirichlet distributed random variable include:

$$\begin{array}{ll} \text{Mean:} & E(x_i) = \frac{\alpha_i}{a} \\ \text{Variance:} & Var(x_i) = \dfrac{\alpha_i (a - \alpha_i)}{a^2 (a+1)} \\ \text{Covariance:} & Cov(x_i, x_j) = -\dfrac{\alpha_i \alpha_j}{a^2 (a+1)}, \end{array}$$

The Dirichlet distribution is often used as a prior pdf for a vector of parameters that have the interpretation of probabilities.

The **Beta distribution** is a special case of the Dirichlet distribution with $N = 2$. Note that with $N = 2$, $x_2 = 1 - x_1$, so this is a univariate distribution for the random variable x_1

(typically just referred to as x). The distribution for x is denoted $x \sim \text{Beta}(\alpha_1, \alpha_2)$, and the pdf for x is:

$$p(x) = \left[\frac{\Gamma(\alpha_1 + \alpha_2)}{\prod_{i=1}^{2} \Gamma(\alpha_i)} \right] x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} \text{ for } 0 < x < 1 \text{ and } \alpha_1, \alpha_2 > 0,$$

Summary statistics for a Beta distributed random variable includes:

$$\begin{aligned} &\text{Mean:} && E(x_i) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \\ &\text{Variance:} && Var(x_i) = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2 (\alpha_1 + \alpha_2 + 1)} \end{aligned}$$

In Matlab, the function betarnd(α_1, α_2, m, n) generates an $m \times n$ matrix of independent $\text{Beta}(\alpha_1, \alpha_2)$ random variables.

The function $\left[\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \right]$ is known as the **Beta function** with arguments α_1 and α_2, and is denoted $B(\alpha_1, \alpha_2)$. Referring to the pdf for the Beta distribution, it is clear that since this pdf must integrate to unity over the possible space for x, the following is true:

$$B(\alpha_1, \alpha_2) = \int_0^1 x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1} dx$$

This is a useful result that we will use later. Another useful result is that when α_1 and α_2 are positive integers:

$$B(\alpha_1, \alpha_2) = \frac{(\alpha_1 - 1)!\,(\alpha_2 - 1)!}{(\alpha_1 + \alpha_2 - 1)!}$$

One particularly interesting special case for the Beta distribution is when $\alpha_1 = 1$ and $\alpha_2 = 1$. This yields the Uniform distribution for x over $[0, 1]$.

3.2 Multinomial Distribution

The **multinomial distribution** is a multivariate extension of the binomial distribution. Consider a vector of discrete random variables $X = (x_1, x_2, \cdots, x_N)'$. X has a multinomial distribution with scalar parameter T and a vector of parameters $p = (p_1, p_2, \cdot, \cdot\cdot, p_N)'$ if the joint pmf for X is given by:

$$\Pr(X) = \begin{cases} \frac{T}{x_1!x_2!\cdots x_N!} p_1^{x_1} \cdots p_N^{x_N} & \text{if } x_i = 0, 1, 2, \cdots, T \text{ and } \sum_{i=1}^{N} x_i = T \\ 0 & \text{otherwise} \end{cases},$$

where $0 \leq p_i \leq 1$, $\sum_{i=1}^{N} p_i = 1$, and T is a positive integer.

The typical notation is: $X \sim M(T, p)$.

Summary statistics for a multinomial distributed random variable include:

Mean:	$E(x_i) = Tp_i$
Variance:	$Var(x_i) = Tp_i(1 - p_i)$
Covariance:	$Cov(x_i, x_j) = -Tp_i p_j$,

The multinomial distribution is typically used in Bayesian econometrics with $T = 1$. In this case, X will be a vector holding $n - 1$ zeros and a single one, and there will be N different possible realizations of X. Each realization of X corresponds to $x_i = 1$, $i = 1, , N$ and $\Pr(X) = \Pr(x_i = 1) = p_i$.

3.3 Multivariate Normal Distribution

Consider a vector of continuous random variables $X = (x_1, x_2, \cdots, x_N)'$. X has a **multivariate normal distribution** with $N \times 1$ vector of location parameters $\mu = (\mu_1, \mu_2, \cdots \mu_N)$

and $N \times N$ matrix of scale parameters Σ if the pdf for X is given by:

$$p(X) = (2\pi)^{-N/2}|\Sigma|^{-1/2}\exp\left(-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right) \quad \text{for} \quad \sigma > 0.$$

The typical notation is: $X \sim N(\mu, \Sigma)$

Summary statistics for normally distributed random variables include:

$$\begin{aligned}
&\text{Mean: } E(x_i) = \mu_i \\
&\text{Variance: } \operatorname{Var}(x_i) = \Sigma_{ii} \\
&\text{Covariance: } \operatorname{Cov}(x_i, x_j) = \Sigma_{ij}
\end{aligned}$$

The multivariate normal distribution is a common choice as prior pdf for conditional mean parameters in econometric models.

Marginal and conditional densities for the multivariate Normal distribution are also distributed normal.

In Matlab, the function mvnrnd(μ,Σ,N,d) generates an $N \times d$ matrix of $N(\mu, \Sigma)$ random variables.

3.4 Multivariate t-Distribution

Consider a vector of continuous random variables $X = (x_1, x_2, \cdots, x_N)'$. X has a **multivariate t-distribution** with $N \times 1$ vector of location parameters μ, $N \times N$ matrix of scale parameters Σ, and shape parameter v if the probability density function for X is given by:

$$p(X) = \frac{v^{v/2}\Gamma\left(\frac{v+N}{2}\right)}{\pi^{N/2}\Gamma\left(\frac{v}{2}\right)}|\Sigma|^{-1/2}\left[v + (X-\mu)'\Sigma^{-1}(X-\mu)\right]^{-(v+N)/2} \quad \text{for} \quad v > 0.$$

The typical notation is: $x \sim t(\mu, \Sigma, v)$. The parameter v is often called the "degrees of freedom parameter."

Summary statistics for t-distributed random variables include:

Mean:	μ for $v > 1$.
Median:	μ
Variance-Covariance Matrix:	$\frac{v}{v-2}\Sigma$ for $v > 2$.

The multivariate t-distribution is sometimes used as prior pdf for conditional mean parameters in econometric models.

In Matlab, the function *mvtrnd* can be used to generate multivariate t-distributed random variables.

Marginal and conditional densities for the multivariate t-distribution are also distributed t.

3.5 Normal-Gamma Distribution

Consider a vector of continuous random variables $X = (x_1, x_2, \cdots, x_N)'$, and a scalar random variable h. X has a **normal-gamma distribution** if the conditional pdf of X given h is $X|h \sim N(\mu, h^{-1}\Sigma)$ and the marginal pdf of h is $h \sim \text{Gamma}(m, v)$, where we use the second definition of the Gamma distribution described above. The joint pdf for X and h is then given by:

$$p(X,h) = (2\pi)^{-N/2}(h)^{N/2}|\Sigma|^{-1/2}\exp\left(-\frac{h}{2}(X-\mu)'\Sigma^{-1}(X-\mu)\right)\frac{1}{\left(\frac{2m}{v}\right)^{v/2}}\frac{1}{\Gamma\left(\frac{v}{2}\right)}h^{\frac{v-2}{2}}\exp\left[-\frac{hv}{2m}\right].$$

The typical notation is: $X \sim NG(\mu, \Sigma, m, v)$.

By definition, $h \sim \text{Gamma}(m, v)$. It can be shown that:

$$X \sim t\left(\mu, m^{-1}\Sigma, v\right)$$

3.6 Wishart Distribution

The **Wishart distribution** is a multivariate generalization of the Gamma distribution. Let H be an $N \times N$ positive definite symmetric random matrix. Also, let A be an $N \times N$ matrix of parameters, and let v be a scalar parameter, where $v > 0$. Then H has a Wishart distribution if the pdf for H is given by:

$$p\left(H\right) = \frac{1}{2^{vN/2}\pi^{N(N-1)/4}\prod_{i=1}^{N}\Gamma\left(\frac{v+1-i}{2}\right)}|H|^{(v-N-1)/2}|A|^{-v/2}\text{exp}\left[-\frac{1}{2}tr\left(A^{-1}H\right)\right]$$

where tr denotes the trace operation.

The typical notation is: $H \sim W\left(v, A\right)$.

Summary statistics for Wishart distributed random variables include:

$$\text{Mean:}E\left(H_{ij}\right) = vA_{ij}$$
$$\text{Variance:}Var\left(H_{ij}\right) = v\left(A_{ij}^{2} + A_{ii}A_{jj}\right)$$
$$\text{Covariance:}Cov\left(H_{ij}, H_{km}\right) = v\left(A_{ik}A_{jm} + A_{im}A_{jk}\right)$$

The Wishart distribution is a common choice as prior pdf for the inverse of a matrix of conditional variance and covariance parameters (known as precision parameters).

In Matlab, the function wishrnd(A, v) generates an $N \times N$ matrix with a $W\left(v, A\right)$ distribution.

When $N = 1$, the Wishart distribution reduces to a Gamma distribution.

Similar to the inverted-Gamma distribution, there is also an inverted-Wishart distribution, which is used as a prior for a matrix of conditional variance and covariance parameters.

4 Kernel of a Probability Distribution Function

A pdf for a vector of random variables X typically has the form $p(X) = Kg(X)$, where $g(X)$ includes no terms that can both be factored out of $g(X)$ and do not depend on X. Here, $g(X)$ is called the **kernel** of the pdf $p(X)$ and K is a numerical constant with respect to X whose role is to ensure that $p(X)$ integrates to unity. This implies:

$$\int_X g(X)\, dX = \frac{1}{K}$$

The constant K is often referred to as the "normalizing constant" of $p(X)$.

Suppose we have some unknown distribution $h(X)$. If we are able to show that $h(X) \propto g(X)$, then we know that $h(X) = p(X)$. In other words, X is distributed according to $p(X)$. This will prove to be very useful.