Common Distributions

MFM Practitioner Module: Risk & Asset Allocation

John Dodson

September 16, 2009
Outline

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded Support

Common Tranforms

Common Mixtures

Non-Parametric Distributions
Below is a taxonomy of common distributions, classified primarily by the topology of their support.

- **finite**
  - Dirac
  - Bernoulli

- **countable**
  - binomial
  - geometric
  - Poisson

- **interval**
  - uniform
  - beta

- **half-line**
  - exponential
  - gamma

- **unbounded**
  - normal
  - Cauchy
  - (Lévy) stable

- **transforms**
  - (generalized) Pareto
  - inverse gamma
  - lognormal

- **mixtures**
  - (Gosset) Student $t$
  - negative-binomial

- **non-parametric**
  - empirical
Finite Support

Let us start our tour by considering two special classes of random variables.

**Bernoulli**
A Bernoulli r.v. is a binary bit. It sample space can be characterized by

- 0/1, T/F, heads/tails, win/lose, up/down

and there are only four possible events (what are they?). The only parameter is the probability or odds ratio.

**Dirac**
The sample space for a Dirac r.v. is \( \mathbb{R} \) in principle, but the only event that has any mass is the singleton set \( \{x_0\} \), where \( x_0 \) is the only parameter. It can be thought of as a degenerate version of a continuous r.v. We write its density as \( f_X(x) = \delta(x - x_0) \), which can be represented as a spike.
Countable Support: Discrete Sample Space

**binomial**

If we think of a Bernoulli r.v. as having sample space \( \{0, 1\} \), the sum of \( n \) Bern\((p)\) r.v.'s is a bin\((n, p)\) and its sample space is \( \{0, 1, \ldots, n\} \). From combinatorics, we know

\[
P\{i\} = \binom{n}{i} p^i (1 - p)^{n-i} \quad \forall i \in \{0, 1, \ldots, n\}
\]

**Stirling’s Approximation**

A useful result from calculus is Stirling’s approximation, which says that \( n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \) for \( n \gg 1 \). In particular,

\[
\binom{n}{i} \approx \frac{1}{\sqrt{2\pi n}} \left(\frac{i}{n}\right)^{-i - \frac{1}{2}} \left(1 - \frac{i}{n}\right)^{-n+i - \frac{1}{2}}
\]

for large \( n, i \).
Common Distributions

John Dodson

Outline
Taxonomy
Finite Support
Countable Support
Interval Support
Half-line Support
Unbounded Support
Common Tranforms
Common Mixtures
Non-Parametric Distributions

Countable Support: Discrete Sample Space

geometric

A geometric r.v. is also related to a sequence of Bernoulli r.v.'s. In terms of the coin toss analogy, it is the length of the “streak” of tails tossed before the next head appears. Again from combinatorics, we know

\[ P\{i\} = p(1 - p)^i \quad \forall i \in \{0, 1, \ldots\} \]

- Notice that the sample space here is countably infinite. One could observe a streak of any length for \(0 < p < 1\).
- Notice also that the process underlying this model is memoryless. The fact that one has already observed a streak of length \(n\) is irrelevant: the only parameter is \(p\), the chance of breaking the streak on the next toss.
- The length of \(k\) streaks is a negative binomial r.v.
Countable Support: Discrete Sample Space

Another notable r.v. whose sample space is the nonnegative integers is the Poisson.

**Poisson**

Consider a binomial r.v. with a very large sample space but a very small probability of occurrence. If we take the limit $n \to \infty$ but we fix $p = \lambda/n$, we get

$$P\{i\} = e^{-\lambda} \frac{\lambda^i}{i!}$$

using Stirling’s Approximation and the limit definition of the exponential function,

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = \sum_{i=0}^{\infty} \frac{x^i}{i!} = e^x \quad \forall x \in \mathbb{R}$$

Analogously, we shall see later that the interval *between* rare events is the limiting case of a geometric r.v.
Interval Support

Now let’s move on the r.v.’s whose sample space is a segment of the real line, traditionally taken to be $[0, 1]$.

**uniform**

The probability of observing a value of a uniform r.v. between $0 \leq a < b \leq 1$ is equal to $b - a$.

- Any particular value between zero and one is equally likely to be observed.
- Imagine a binary decimal, e.g. $0.0110\ldots$, where each bit to the right of the decimal place is Bern($\frac{1}{2}$). This is a uniform r.v.
- All modern computer systems can generate an almost-endless stream of uniform (pseudo)-random variates, which can be used for generating samples of other r.v.’s using transformation techniques.
Interval Support

beta

A beta family is parameterized by two continuous parameters, $\alpha, \beta > 0$. The scale factor for the density involves the Gamma function.

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \forall x \in [0, 1]$$

There is a deep connection between the beta and binomial. The formulæ for the densities are essentially the same; but instead of describing the count, the beta describes the probability.

- The beta is a good model for an unknown probability.
- The beta is also a good model for an unknown fraction.
- The uniform is a special case, $\text{beta}(1, 1) \sim \text{U}([0, 1])$. 

Half-line Support

**exponential**

The limiting case of the geometric with $p = \lambda/n$ and $n \to \infty$ is the exponential. Its CDF is simply

$$F_X(x) = 1 - e^{-\lambda x} \quad \forall x \geq 0$$

Differentiating, we get the density.

$$f_X(x) = \lambda e^{-\lambda x} \quad \forall x \geq 0$$

- Note that the mode of an exponential r.v. is zero.
- The minimum of $n \exp(\lambda)$ r.v.'s is $\exp(n\lambda)$
- The interval between arrivals of a Poisson process is exponential.
Half-line Support

One can arrive at the gamma family by several disparate tracks. Here, we will approach it as a sum of exponentials.

**gamma**

Since the characteristic function of $Y \sim \exp(\lambda)$ is
\[
\phi_Y(t) = (1 - \frac{it}{\lambda})^{-1}
\]
(prove), the characteristic function of the sum of $k$ exponential r.v.'s is
\[
\phi_{Y_1 + \ldots + Y_k}(t) = (1 - \frac{it}{\lambda})^{-k}.
\]
We can apply the Fourier transform to get the density of $X = Y_1 + \ldots + Y_k$,
\[
f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \quad \forall x \geq 0
\]

Later, we will see that this is also the density of a sum of squared normals\(^1\), and even a natural description for the space of random positive-definite matrices.

\(^1\)which is how Meucci introduces it.
Unbounded Support

We have already talked about the normal, whose reputation benefits from the Central Limit Theorem. But not every r.v. has a finite variance. The simplest example of an unbounded r.v. without a finite variance is the Cauchy.

Cauchy

The standard version\(^2\) of the Cauchy has the density

\[ f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \]

whose graph looks like the density of a normal; but statistically it is nothing like a normal.

- The standard Cauchy has the characteristic function
  \[ \phi_X(t) = e^{-|t|}. \]
- The ratio of two normal r.v.’s is a Cauchy.

\(^2\)Any affine transformation of a Cauchy is still a Cauchy.
stable

The Cauchy is a special case of the exotic (Lévy) stable family. Not only do stable r.v.’s lack a variance, but (excepting the Cauchy) they also lack a tractable density. The standardized stable characteristic function is

$$
\phi_X(t) = \begin{cases} 
    e^{-|t|^\alpha(1-\text{sgn}(t)i\beta \tan(\alpha\pi/2))} & \alpha \neq 1 \\
    e^{-|t|(1+\text{sgn}(t)i\beta(2/\pi) \log |t|)} & \alpha = 1
\end{cases}
$$

for parameters $0 < \alpha < 2$ and $-1 < \beta < 1$.

- The Cauchy corresponds to $\alpha = 1$ and $\beta = 0$
- $\beta \geq 0$ allows for the density to be asymmetric.
- The limit $\alpha \to 2$ is $N(0, 2)$. The limit $\alpha \to 0$ is $\delta(0)$. 
Pareto
A basis for the concept of power laws is a result about the frequency of extreme events. This has at its heart a (generalized) Pareto r.v. which is an exponential of an exponential, \( \log (1 + \xi X) \sim \exp (1/\xi) \).

The resulting distribution function is

\[
F_X(x) = 1 - (1 + \xi x)^{-1/\xi}
\]

\( \forall \begin{cases} 0 \leq x & \xi \geq 0 \\ 0 \leq x \leq -\frac{1}{\xi} & \xi < 0 \end{cases} \)

- \( \xi \) is sometimes called the tail index
- note that if \( U \sim U([0, 1]) \), then \( \frac{U^{-\xi} - 1}{\xi} \sim \text{Pareto}(\xi) \)
The gamma is a reasonable model for an unknown scale factor. But sometimes your application will call for dividing rather than multiplying.

**inverse gamma**

It is a simple matter to evaluate the density of the reciprocal of a gamma r.v.,

\[
f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{1-k} e^{-\lambda/x} \quad \forall x > 0
\]

for parameters \( \lambda > 0 \) and \( k > 0 \).

- the inverse gamma comes up in Bayesian analysis
- we will see that it is a natural description for an unknown variance
Common Transforms

Since most assets cannot become liabilities, the support for future asset values (or prices) is $\mathbb{R}^+$ (or some subset). Furthermore, since holdings values are usually price $\times$ quantity, nominal prices are rarely important. Hence a natural model for the future value of an asset is as a lognormal r.v.

**lognormal**

The logarithm of a lognormal r.v. is normal. It has a density

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{e^{-\frac{\log(x/\mu)^2}{2\sigma^2}}}{x} \quad \forall x > 0$$

for parameters $\mu > 0$ and $\sigma > 0$.

- Note that, unlike the normal, the expected value of a lognormal r.v. involves both parameters: $E X = \mu e^{\frac{1}{2} \sigma^2}$. 
The Student r.v. is not traditionally introduced as a mixture, but this interpretation will be useful to us later.

*(Gosset)* Student $t$

Consider a normal r.v. with an unknown variance close to one. If the variance is a draw from an inverse gamma, 

$$X \mid \sigma^2 \sim N(0, \sigma^2)$$

$$\sigma^2 \sim \text{Gamma}^{-1} \left( \frac{\nu}{2}, \frac{\nu}{2} \right)$$

we can write down the resulting unconditioned density.

$$f_X(x) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma \left( \frac{\nu}{2} \right) \Gamma \left( \frac{1}{2} \right) \sqrt{\nu}} \left( 1 + \frac{x^2}{\nu} \right)^{-\left(\frac{\nu+1}{2}\right)}$$

For historical reasons, if the parameter $\nu > 0$ is an integer, the r.v. is terms to have $\nu$ degrees of freedom.
Common Mixtures

If you are working with an r.v. that is Poisson, but you only have an estimate for the parameter, one approach is to say that the true parameter is a draw of a gamma r.v. which you can confidently characterize.

**negative binomial**

This is called the gamma-Poisson mixture, and the result is called the negative binomial. The hierarchical model is

\[ X \mid \lambda \sim \text{Poisson}(\lambda) \]

\[ \lambda \sim \text{Gamma} \left( k, \frac{p}{1 - p} \right) \]

with the result

\[ P_X \{ i \} = \frac{\Gamma(i + k)}{i! \Gamma(k)} p^k (1 - p)^i \quad \forall i \in \{0, 1, \ldots\} \]
Non-Parametric Distributions

A modern trend in statistics is to move away from parametric descriptions towards non-parametric descriptions.

**empirical**

The most natural non-parametric description of a r.v. $X$ based on a dataset $\{x_1, x_2, \ldots, x_n\}$ is the empirical r.v.

\[
    f_X(x) = \frac{1}{n} \sum_{i=1}^{n} \delta (x - x_i)
\]

\[
    F_X(x) = \frac{1}{n} \sum_{i=1}^{n} H(x - x_i)
\]

- This can be regularized by replacing the Dirac deltas by normal densities with sufficiently small variances. This is also termed kernel smoothing.