

Probability Distributions

MFM Practitioner Module: Quantitative Risk Management

John Dodson

September 23, 2015

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
Support

Common
Transforms

Common Mixtures

Non-Parametric
Distributions

Multivariate
Distributions

Outline

Probability
Distributions

John Dodson

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded Support

Common Transforms

Common Mixtures

Non-Parametric Distributions

Multivariate Distributions

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
Support

Common
Transforms

Common Mixtures

Non-Parametric
Distributions

Multivariate
Distributions

Definitions

For our purposes, A **random variable** is a quantity whose value is not known to us right now (but may be at some point in the future). These can be represented mathematically as **measurable functions** of a **sample space**, and we will usually denote them by upper-case Roman letters. We will usually denote a particular value obtained by a random variable by the corresponding lower-case letter. There must be a probability associated with every possible set of outcomes. Such sets are called **events** and might consist of intervals or points or any combination thereof. The corresponding probability is called the **probability measure** of the event.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
Support

Common
Transforms

Common Mixtures

Non-Parametric
Distributions

Multivariate
Distributions

This structure lends itself to a measure theory interpretation, where the probability associated with a set is simply the integral of the **probability density** over that set.

$$P \{X \in (a, b)\} = \int_a^b f_X(x) dx$$

If X ranges over the real numbers, \mathbb{R} , then $f_X(\cdot)$ must have certain properties. In particular, it must be a non-negative (generalized) function and

$$\lim_{x \rightarrow -\infty} f_X(x) = 0 \quad \lim_{x \rightarrow \infty} f_X(x) = 0$$
$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

In addition to the **density function**, there are at least three other equivalent characterizations of a real random variable

- ▶ **distribution function** $F_X(x) = \int_{-\infty}^x f_X(x') dx'$
- ▶ **quantile function** $q_p(F_X) = F_X^{\leftarrow}(p)$
- ▶ **characteristic function** $\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$

The quantile function is defined in terms of the **generalized inverse**

$$F^{\leftarrow}(p) = \inf \{x : F(x) \geq p\}$$

The last is based on the Fourier transform, where $i^2 \triangleq -1$.

Taxonomy of Random Variables

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's) stable
- ▶ *transforms*
 - ▶ generalized Pareto
 - ▶ gen. inv. Gaussian
 - ▶ lognormal
- ▶ *mixtures*
 - ▶ (Gosset's) Student's- t
 - ▶ negative-binomial
 - ▶ generalized hyperbolic
- ▶ *non-parametric*
 - ▶ empirical

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
Support

Common
Transforms

Common Mixtures

Non-Parametric
Distributions

Multivariate
Distributions

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

Bernoulli

A Bernoulli r.v. is a “bit”. Its 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?**). It has only one **parameter** (not necessarily equal to $\frac{1}{2}$).

Dirac

The sample space for a Dirac r.v. is, in principle, \mathbb{R} ; 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 the degenerate continuous r.v. We write its density as $f_X(x) = \delta(x - x_0)$, which can be represented as a **spike**.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

Countable Support: Discrete Sample Space

binomial

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

$$P\{i\} = \binom{n}{i} p^i (1-p)^{n-i} \quad \forall i \in \{0, 1, \dots, n\} \subset \mathbb{Z}$$

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 .

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
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

$$P\{i\} = p(1 - p)^i \quad \forall i \in \{0, 1, \dots\}$$

- ▶ 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 $n \rightarrow \infty$ 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,

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

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

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. $0.0110\dots$, where each bit to the right of the decimal place is Bern $\left(\frac{1}{2}\right)$. 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.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

beta

A beta r.v. 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, $\text{beta}(1, 1) \sim U([0, 1])$.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

exponential

The limiting case of the geometric with $p = \lambda/n$ and $n \rightarrow \infty$ is the exponential. Its distribution function 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.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

Half-line Support

One can arrive at a Gamma r.v. by several routes. 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

$\phi_{Y_1+\dots+Y_k}(t) = (1 - \frac{it}{\lambda})^{-k}$. We can apply the Fourier transform to get the density of $X = Y_1 + \dots + Y_k$,

$$f_X(x) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \quad \forall x \geq 0$$

- ▶ This is also a sum of squared normals (**Chi-squared**)
- ▶ and a natural description for the space of random positive-definite matrices (**Wishart**)

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

Unbounded Support

The most important distribution for $X \in \mathbb{R}$ is the **normal** or **Gaussian** distribution, $X \sim \mathcal{N}(\mu, \sigma^2)$ with μ the **mean** and σ^2 the **variance** (σ is the **standard deviation**).

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$F_X(x) \triangleq \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)$$

$$q_p(F_X) = \mu + \sigma \Phi^{-1}(p) = \mu + \sigma\sqrt{2} \operatorname{erf}^{-1}(2p - 1)$$

$$\phi_X(t) = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$$



Normal density

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

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 version¹ of the Cauchy has the density

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

whose graph looks like that 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.

¹Any affine transformation of a Cauchy is still a Cauchy.

(Lévy's) stable

The Cauchy is a special case of the (Lévy's) 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 $\alpha \rightarrow 2$ is $\mathcal{N}(0, 2)$. The limit $\alpha \rightarrow 0$ is $\delta(0)$.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

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

$$F_X(x) = 1 - (1 + \xi x)^{-1/\xi} \quad \forall \begin{cases} 0 \leq x & \xi \geq 0 \\ 0 \leq x \leq -\frac{1}{\xi} & \xi < 0 \end{cases}$$

- ▶ ξ is sometimes called the **tail index**
- ▶ note that if $U \sim U(0, 1)$, then $\frac{U^{-\xi}-1}{\xi} \sim \text{Pareto}(\xi)$

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

The gamma is a reasonable model for an unknown scale factor. But sometimes your application will call for dividing rather than multiplying.

Reciprocal 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 reciprocal gamma comes up in Bayesian analysis
- ▶ it is a natural description for an unknown variance

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

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: $E X = \mu e^{\frac{1}{2}\sigma^2}$.

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

The gamma and reciprocal gamma can be combined into a generalized family for r.v.'s with half-line support.

Generalized Inverse Gaussian

$$f_X(x) = \frac{\chi^{-\lambda} (\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} e^{-\frac{\chi}{2x} - \frac{\psi x}{2}}$$

for $x > 0$, where $K(\cdot)$ is Bessel's function of the third kind.

- ▶ There are several versions of parameterization in use
- ▶ Other members of this family include the **inverse Gaussian** and the **reciprocal inverse Gaussian**
- ▶ The name relates to the first passage time of a Brownian motion through a boundary

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

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

$$X|\lambda \sim \text{Poisson}(\lambda)$$

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

with the result

$$P_X\{i\} = \frac{\Gamma(i+k)}{i!\Gamma(k)} p^k (1-p)^i \quad \forall i \in \{0, 1, \dots\}$$

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

Student's- t is a symmetric r.v. which exhibits leptokurtosis.

(Gosset's) Student's- t

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

$$\begin{aligned}X \mid \sigma^2 &\sim \mathcal{N}(0, \sigma^2) \\ \sigma^2 &\sim \text{Gamma}^{-1}\left(\frac{\nu}{2}, \frac{\nu}{2}\right)\end{aligned}$$

the resulting unconditioned density is

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

- ▶ $\nu = 1$ is a Cauchy
- ▶ The limit $\nu \rightarrow \infty$ is a normal

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

The version of the Student's- t above has a variance for $\nu > 2$, but it is not unity.

Standardized Student's- t

The standardized version is useful for fitting residuals. It has the density

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

- ▶ Note that $E e^X \rightarrow \infty$ for any finite ν . *This would seem to be a problem for a model of log-returns!*

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

The Student's- t is part of the **generalized hyperbolic** family, which is based on a normal mean-generalized inverse Gaussian variance mixture (with $\lambda = -\nu/2$).

Normal-reciprocal inverse Gaussian (NRIG)

The symmetric generalized hyperbolic r.v. with $\lambda = \frac{1}{2}$, is particularly useful. The standardized version has the density

$$f_X(x) = \frac{1}{\pi} e^{gx} \sqrt{1+g} K_0 \left(\sqrt{g^2 + (1+g)x^2} \right)$$

in the Babbs representation for shape parameter $g > 0$. It is not obvious, but the limit $g \rightarrow \infty$ corresponds to the normal.

- ▶ As a model for the residuals of the log-returns for asset prices, this is superior to the Student's- t example from the text because $E e^X$ is finite!

Outline

Random Variables

Characterizations

Taxonomy

Finite Support

Countable Support

Interval Support

Half-line Support

Unbounded
SupportCommon
Transforms

Common Mixtures

Non-Parametric
DistributionsMultivariate
Distributions

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 $\{x_1, x_2, \dots, x_n\}$ 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, which is termed **kernel smoothing**.

Multivariate Distributions

Many of the univariate distributions can be generalized to a vector-valued sample space

- ▶ stable family and related, including the normal, Cauchy, Student's- t , and generalized hyperbolic
- ▶ discrete and empirical
- ▶ uniform

For the uniform, this is just a matter of defining the support in \mathbb{R}^n and normalizing appropriately. For the others, often the replacement

$$\left(\frac{X - \mu}{\sigma}\right)^2 \rightarrow (X - \mu)' \Sigma^{-1} (X - \mu)$$

and **normalization** is all that is required to lift the sample space from \mathbb{R} to \mathbb{R}^N .

- ▶ A multivariate density can always be used as a basis for a copula.