Intermediate Probability

Alex Houtz*

August 2, 2023

1 Moment-Generating and Characteristic Functions

Sometimes, it is useful in proofs to have a function that characterizes distributions. The first function we will look at is the **moment generating function** (MGF). The MGF of X is defined as:

$$M(t) = \mathbb{E}\left[e^{tX}\right]$$

We find the moments of the distribution X by taking the derivative of $M(t)$ with respect to t and then setting t to zero.

Example 1. Let $X \sim N\left(\mu, \sigma^2\right)$. Then the MGF is:

$$\begin{aligned}
\mathbb{E}\left[e^{tx}\right] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2-2\mu x+\mu^2)+tx} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2-2\mu x+\mu^2-2\sigma^2 tx)} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2-2(\mu+\sigma^2 t)x+\mu^2)} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\mu^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2-2(\mu+\sigma^2 t)x)} dx
\end{aligned}$$

Now we complete the square inside the exponential that is still within the integral:

*Math Camp Instructor | University of Notre Dame

$$(x - (\mu + \sigma^2 t))^2 = x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2$$

We are missing the last term, so multiplying and dividing by the appropriate terms gives:

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\mu^2} e^{\frac{1}{2\sigma^2}(\mu+\sigma^2 t)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x^2 - 2(\mu+\sigma^2 t)x + (\mu+\sigma^2 t)^2)} dx \\
&= e^{-\frac{1}{2\sigma^2}\mu^2} e^{\frac{1}{2\sigma^2}(\mu+\sigma^2 t)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-(\mu+\sigma^2 t))^2} dx
\end{aligned}
$$

Note that the term inside the integral is just a random variable with distribution $N(\mu + \sigma^2 t, \sigma^2)$, which by definition integrates to one. Simplifying the remaining expression gives us the moment generating function for a normal distribution:

$$M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Let's find the first moment, $\mathbb{E}[x]$:

$$\frac{\partial}{\partial t} M(t)\bigg|_{t=0} = (\mu + \sigma^2 t) e^{\mu t + \frac{1}{2}\sigma^2 t^2}\bigg|_{t=0} = \mu$$

Sometimes, a random variable does not have a MGF. In contrast, every (real-valued) random variable has a **characteristic function** that completely defines its probability distribution. If a random variable has a pdf, then the characteristic function is the Fourier transform of that pdf. Formally:

$$\varphi(t) = \mathbb{E}[e^{itX}]$$

where $i = \sqrt{-1}$. Of course, if we take the transform $\varphi(-it)$, the characteristic function equals the MGF. Following the steps in the example above, we can easily derive the characteristic function of the normal distribution:

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

The k^{th} moment is derived as follows:

$$\varphi^{(k)}(0) = i^k \mathbb{E}[X^k]$$

where $\varphi^{(k)}$ is the k^{th} derivative of the characteristic function.

2 Multivariate Probability

2.1 Joint Distributions

Suppose we have two random variables, X and Y. We can define a **joint** distribution over X and Y. Now, though, we have two supports, $\mathcal{X}$ and $\mathcal{Y}$. What if we want to find the pdf for just X? We call this the **marginal** distribution. We simply integrate out Y:

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

To make this more concrete, let's look at an example:

Example 2. Let X and Y have joint density $f(x, y) = e^{-x-y}$ for $x, y > 0$. Find $f_x(x)$ and $f_y(y)$. Determine if $X \perp\!\!\!\perp Y$.

Let's first integrate out Y:

$$\begin{aligned} f_x(x) &= \int_0^{\infty} e^{-x-y} dy \\ &= \left[-e^{-x-y}\right]_0^{\infty} \\ &= 0 + e^{-x} \\ &= e^{-x} \end{aligned}$$

We can see that the problem is symmetric, so $f_y(y) = e^{-y}$. Recall that two random variables are independent if $f(x)f(y) = f(x, y)$. Then:

$$\begin{aligned} f_x(x) f_y(x) &= e^{-x} e^{-y} \\ &= e^{-x-y} \\ &= f(x, y) \end{aligned}$$

So X and Y are indeed independent.

Once we find the marginals, we can take expectations, variances, etc. as normal.

We can also do a multivariate change-of-variables. The formula is essentially the same as the univariate case, but doing the actual math can be a little complicated:

Example 3. Using the joint pdf from the previous problem ($f(x, y) = e^{-x-y}$), apply the two transformations $Z = X - Y$ and $W = X + Y$. Find the marginal distributions for Z and W.

Solution: Finding the inverse functions

We begin with solving for X. Rearranging the transformation equations, we find that $Z + Y = X$ and $W - Y = X$. Setting these equal:

$$\begin{aligned} Z + Y &= W - Y \\ 2Y &= W - Z \\ Y &= \frac{W - Z}{2} \end{aligned}$$

Similarly, by solving for Y, we find that:

$$\begin{aligned} -Z + X &= W - X \\ 2X &= W + Z \\ X &= \frac{W + Z}{2} \end{aligned}$$

Solution: Plug inverse functions into the joint pdf

$$\begin{aligned} f_{xy}(x, y) &= e^{-x-y} \\ f(g^{-1}(z, w), h^{-1}(z, w)) &= e^{\frac{-w-z}{2} + \frac{z-w}{2}} \\ &= e^{\frac{-2w}{2}} \\ &= e^{-w} \end{aligned}$$

Solution: The Jacobian

The Jacobian is:

$$
\begin{aligned}
|J| &= \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} \\
&= \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} \\
&= \frac{1}{4} + \frac{1}{4} \\
&= \frac{1}{2}
\end{aligned}
$$

Solution: Substitute into the change of variables formula

Using the same formula as before

$$
\begin{aligned}
f_{zw}(z, w) &= e^{-w} \cdot \frac{1}{2} \\
&= \frac{e^{-w}}{2}
\end{aligned}
$$

Now we have our joint pdf!

Solution: Finding the marginal pdfs

First we need to figure out the support for W and Z. We know that:

$$
\begin{aligned}
0 < x < \infty \qquad & \qquad 0 < y < \infty \\
0 < \frac{w+z}{2} < \infty \qquad & \qquad 0 < \frac{w-z}{2} < \infty \\
0 < w + z < \infty \qquad & \qquad 0 < w - z < \infty
\end{aligned}
$$

From here we can determine that

$$
-w < z < w \qquad\qquad 0 < w < \infty
$$

We are ready to find the marginal distribution for W:

$$
\begin{aligned}
f_w(w) &= \int_{-w}^{w} \frac{e^{-w}}{2} dz \\
&= \left[\frac{1}{2} e^{-w} z\right]_{-w}^{w} \\
&= \frac{1}{2} e^{-w} w + \frac{1}{2} e^{-w} w \\
&= w e^{-w} \qquad\qquad \text{for } 0 < w < \infty
\end{aligned}
$$

One more to go! We need to re-evaluate the support before we can continue. We can write the support as:

$$
-z < w < \infty \qquad z < w < \infty \qquad\qquad \text{with } -\infty < z < \infty
$$

So:

$$
\begin{aligned}
f_z(z) &= \begin{cases} \int_{-z}^{\infty} \frac{e^{-w}}{2} dw & \text{for } -z < w < \infty \\ \int_{z}^{\infty} \frac{e^{-w}}{2} dw & \text{for } z < w < \infty \end{cases} \\
&= \begin{cases} \left[\frac{-1}{2} e^{-w}\right]_{-z}^{\infty} & \text{for } z < 0 \\ \left[\frac{-1}{2} e^{-w}\right]_{z}^{\infty} & \text{for } z > 0 \end{cases} \\
&= \begin{cases} \frac{1}{2} e^{z} & \text{for } z < 0 \\ \frac{1}{2} e^{-z} & \text{for } z > 0 \end{cases}
\end{aligned}
$$

If we combine these two cases, we arrive at our answer:

$$
f_z(z) = \frac{1}{2} e^{-|z|}
$$

This is the Laplace distribution's pdf with a location parameter of zero and a scale parameter of 1.

2.2 Covariance and Correlation

Now that we have two random variables, we can calculate the covariance between X and Y:

$$Cov(x, y) = \mathbb{E}\left[(x - \mathbb{E}[x])(y - \mathbb{E}[y])\right]$$

Like variance, we can simplify the covariance formula:

$$\begin{aligned}
\mathbb{E}\left[(x - \mathbb{E}[x])(y - \mathbb{E}[y])\right] &= \mathbb{E}\left[xy - x\mathbb{E}[y] - \mathbb{E}[x]y + \mathbb{E}[x]\mathbb{E}[y]\right] \\
&= \mathbb{E}[xy] - \mathbb{E}[x][y] - \mathbb{E}[x][y] + \mathbb{E}[x][y] \\
&= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]
\end{aligned}$$

Note that if $X \perp\!\!\!\perp Y$, the covariance is zero, as we can separate the expected values:

$$\begin{aligned}
Cov(x, y) &= \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] \\
&= \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[x]\mathbb{E}[y] \\
&= 0
\end{aligned}$$

But the converse is not necessarily true. Let $X \sim Unif(-1, 1)$ and $Y = X^2$. Then:

$$\begin{aligned}
Cov(x, y) &= Cov(x, x^2) \\
&= \mathbb{E}[xx^2] - \mathbb{E}[x]\mathbb{E}[x^2] \\
&= \mathbb{E}[x^3] - \mathbb{E}[x]\mathbb{E}[x^2] \\
&= 0
\end{aligned}$$

But clearly X and Y are dependent.

To normalize how correlated two random variables are, we calculated the Pearson correlation coefficient, often simply referred to as the **correlation**, denoted by $\rho_{x,y}$. We calculate the correlation as:

$$\rho_{x,y} = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

where σ_x and σ_y are the standard deviations of X and Y, respectively. $\rho_{x,y}$ will always be between -1 and 1.

We can use the linear property of expectations for multivariate linear combinations as well. Suppose we have two random variables, X and Y, and three constants: a, b, and c. Let:

$$Z = a + bX + cY$$

What is $\mathbb{E}[Z]$? What is $Var(Z)$?

This is a linear equation, so the expected value is simple:

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[a + bX + cY] \\ &= a + b\mathbb{E}[X] + c\mathbb{E}[Y]\end{aligned}$$

The variance is now a little more complex:

$$\begin{aligned}Var(Z) &= \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \\ &= \mathbb{E}[(a + bX + cY)^2] - (a + b\mathbb{E}[X] + c\mathbb{E}[Y])^2 \\ &= \mathbb{E}[a^2 + 2abX + 2acY + b^2X^2 + 2bcXY + c^2Y^2] - a^2 - 2ab\mathbb{E}[X] \\ &\qquad - 2ac\mathbb{E}[Y] - b^2\mathbb{E}[X]^2 - 2bc\mathbb{E}[X]\mathbb{E}[Y] - c^2\mathbb{E}[Y]^2 \\ &= b^2\left(\mathbb{E}[X^2] - \mathbb{E}[X]^2\right) + c^2\left(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2\right) + 2bc\left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right) \\ &= b^2Var(X) + c^2Var(Y) + 2bcCov(X, Y)\end{aligned}$$

Theorem 1. *Suppose that X and Y are independently distributed $N(\mu_x, \sigma_x^2)$ and $N(\mu_y, \sigma_y^2)$ respectively. Then the sum of X and Y will be normally distributed:*

$$X + Y \sim N\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2\right)$$

If X and Y are jointly normally distributed, but not independent, then the sum of X and Y is:

$$X + Y \sim N\left(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}\right)$$

I will prove the first part, as the second part gets messy quickly:

Proof. Take the characteristic function of a normal distribution:

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

Because X and Y are independent, then the joint distribution is just the product of the marginals. Thus, the joint characteristic function is just the product of the individual characteristic functions:

$$\begin{aligned}\varphi_{x+y}(t) &= \varphi_x(t)\varphi_y(t) \\ &= e^{it\mu_x - \frac{1}{2}\sigma_x^2 t^2} e^{it\mu_y - \frac{1}{2}\sigma_y^2 t^2} \\ &= e^{it(\mu_x+\mu_y) - \frac{1}{2}(\sigma_x^2+\sigma_y^2)t^2}\end{aligned}$$

But this is just the characteristic function of a normal distribution with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$. ■

2.3 Conditional Distributions

Just like with probability functions, we also have conditional distributions:

$$f_{y|x}(y|x) = \frac{f(x, y)}{f_x(x)}$$

To take the expectation of the conditional distribution:

$$\mathbb{E}\left[Y|X\right] = \int_{-\infty}^{\infty} y f_{y|x}(y|x) dy$$

Sometimes, we want to take the pdf of a variable conditioning on that variable being less than some value, b. That pdf is:

$$f(x|X \leq b) = \frac{f(x)}{F(b)}$$

where the denominator is simply the CDF evaluated at $F(b) - F(0)$. The expected value is then:

$$\mathbb{E}[x|X \leq b] = \int_{-\infty}^{b} x \frac{f(x)}{F(b)} dx$$

Conditional distributions lead to a very powerful theorem:

Theorem 2 (Law of Iterated Expectations)**.**

$$\mathbb{E}[Y] = \mathbb{E}_x[\mathbb{E}[Y|X]]$$

Proof. Start with the definition of the expectation and proceed through the calculus:

$$\begin{aligned}
\mathbb{E}_x[\mathbb{E}[Y|X]] &= \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_x(x) dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{y|x}(y|x) f_x(x) dy dx \\
&= \int_{\infty}^{\infty} y f(y, x) dy dx \\
&= \mathbb{E}[Y]
\end{aligned}$$

which is the law of iterated expectations. ■

Similarly, there is a theorem for variance:

Theorem 3 (Law of Total Variance)**.**

$$Var(Y) = \mathbb{E}\left[Var(Y|X)\right] + Var\left(\mathbb{E}[Y|X]\right)$$

Proof. Start with the variance decomposition formula and use the law of iterated expectations:

$$\begin{aligned}
Var(Y) &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\
&= \mathbb{E}\left[\mathbb{E}[Y^2|X]\right] - \mathbb{E}\left[\mathbb{E}[Y|X]\right]^2
\end{aligned}$$

Substitute in the variance decomposition formula for the first term:

$$Var(Y) = \mathbb{E}\left[Var(Y|X) + \mathbb{E}[Y|X]^2\right] - \mathbb{E}\left[\mathbb{E}[Y|X]\right]^2$$

Regroup terms:

$$Var(Y) = \mathbb{E}\left[Var(Y|X)\right] + \left(\mathbb{E}\left[\mathbb{E}[Y|X]^2\right] - \mathbb{E}\left[\mathbb{E}[Y|X]\right]^2\right)$$

Recognizing $Var(Y|X)$ gives us:

$$Var(Y) = \mathbb{E}\left[Var(Y|X)\right] + Var\left(\mathbb{E}[Y|X]\right)$$

which is the law of total variance. ■

2.4 Vectors of Random Variables

So far, we have only looked at bivariate distributions. But we could have as many variables together as we want. Let X be an $n \times 1$ random vector:

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}'$$

All of the concepts we have covered so far extend naturally to vectors of random variables. For example, take the expectation of the random vector:

$$\mathbb{E}[X] = \begin{bmatrix} \mathbb{E}[x_1] & \mathbb{E}[x_2] & \dots & \mathbb{E}[x_n] \end{bmatrix}'$$

Importantly, each x_i could have a different marginal distribution, so the expectations are over each x_i's distribution. We can also take the variance over the vector of random variables:

$$Var(X) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \dots & \sigma_n^2 \end{bmatrix}$$

We commonly denote this matrix as Σ. You may have heard this called the **variance-covariance**, or just covariance, matrix. Importantly, Σ is always a symmetric matrix, and, in practice, almost always positive definite. If a matrix is positive-definite, it has a **Cholesky decomposition**:

$$\Sigma = \Sigma^{1/2}\Sigma^{(1/2)'}$$

Where $\Sigma^{1/2}$ is the lower triangular matrix and $\Sigma^{(1/2)'}$ is the upper triangular matrix (more on this tomorrow). We use Cholesky decompositions of the covariance matrix when identifying shocks to macro systems. It can also be used to quickly solve systems of linear equations.

What if we want to linearly transform a random vector? For example define Y as:

$$Y = AX$$

where A is some matrix such that matrix multiplication is defined. The expectation works as usual:

$$\mathbb{E}[Y] = A\mathbb{E}[X]$$

But the variance will look slightly different. Recall that with a constant a, $Var(aX) = a^2 Var(X)$. Similarly, we want to "square" the matrix A. To do so, we just post-multiply by A':

$$\begin{aligned} Var(Y) &= Var(AX) \\ &= AVar(X)A' \\ &= A\Sigma_x A' \end{aligned}$$

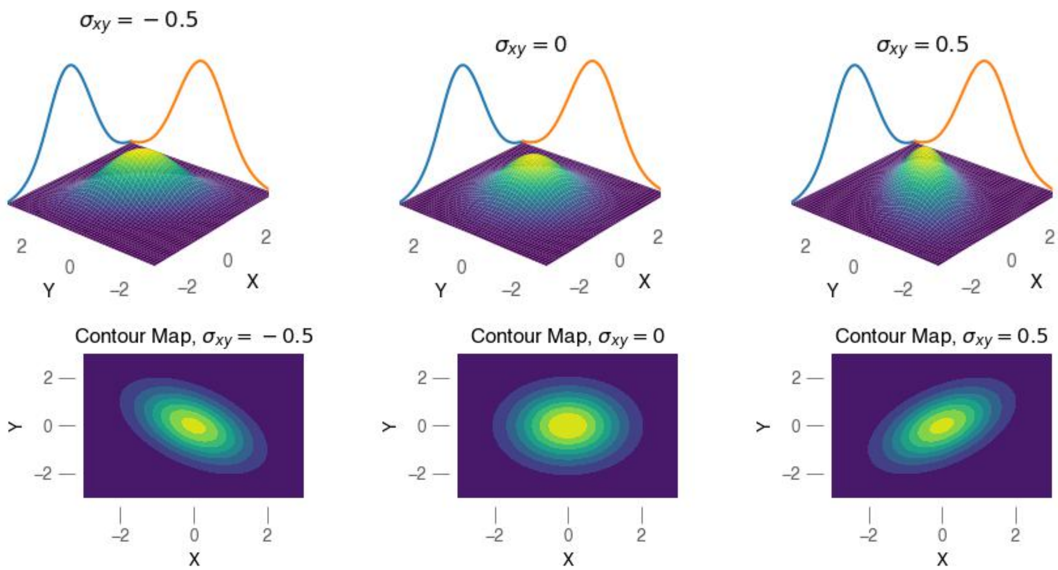
Multivariate Normal

Probably the most important distribution is the normal distribution. Like the univariate normal, we can write the pdf for a multivariate normal. Let X be an $n \times 1$ vector of random variables. Then if $X \sim N(\mu, \Sigma)$, it has pdf:

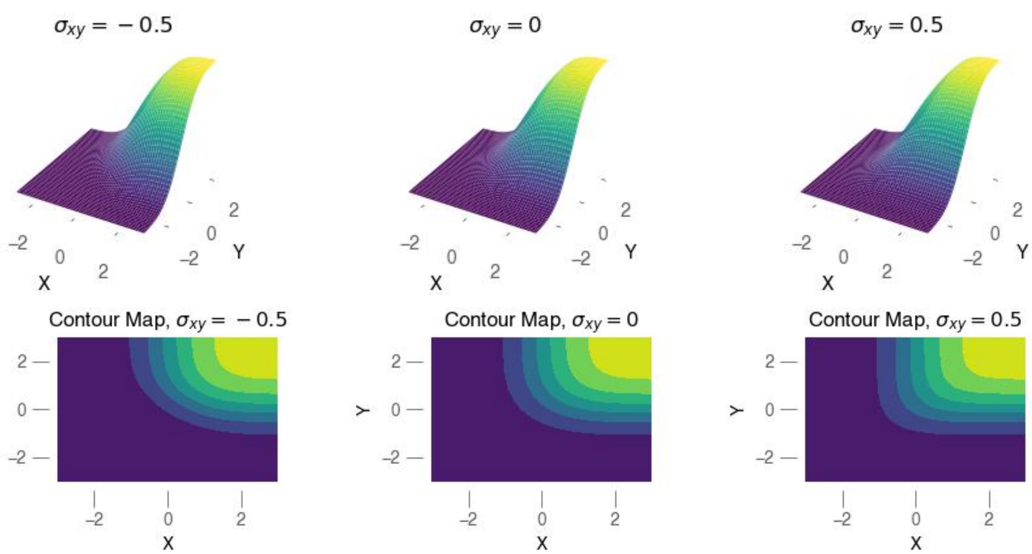
$$f(x|\mu, \Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^n |\Sigma|^{-1/2} \, e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

The pdf and contour plots for the bivariate case are shown in the figure below. We can see that the covariance values matter for the shape of the distribution. But what about for the CDFs? The CDFs themselves all look pretty similar, but when we take the contours we can see the difference in the CDFs. Because the CDF is a sum of the joint probabilities over both X and Y, the positively-correlated bivariate normal begins to accumulate density before the others. This makes sense, as in the pdf contour plots density is higher in the lower left corner for the positively-correlated distribution than for the other two.

Bivariate Normal Distribution



Bivariate Normal Distribution - CDFs

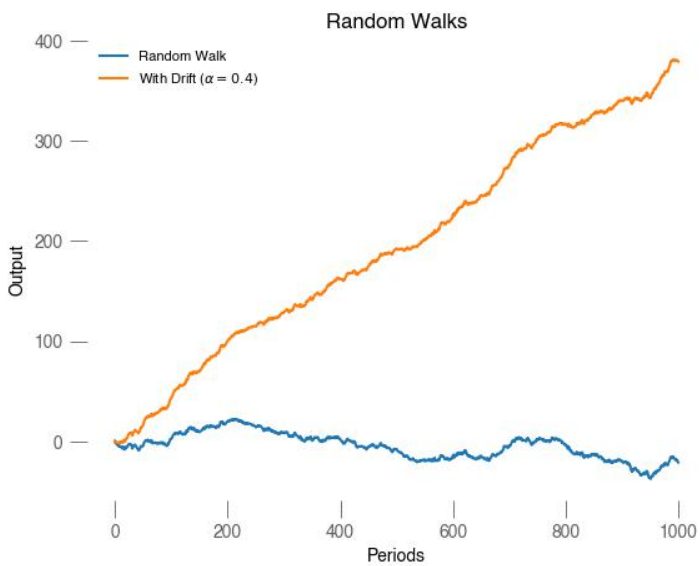


3 Autoregressive Variables

An **autoregressive** variable is a variable that's value is dependent on its previous values. This area of probability is massive, so we will just briefly cover AR(p) processes, where p refers to the number of lags in the equation. Specifically, let's focus on an AR(1):

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t$$

where α is a constant, ρ is a persistence term, and $\varepsilon_t \sim N(0, \sigma^2)$ and is *iid* across time. There are a few combinations of these α and ρ that give common models. First, consider $\rho = 1$ and $\alpha = 0$. Then we have a **random walk**. Simulating a random walk gives the blue line in the plot below. If I add an $\alpha \neq 0$, we get a random walk with drift.



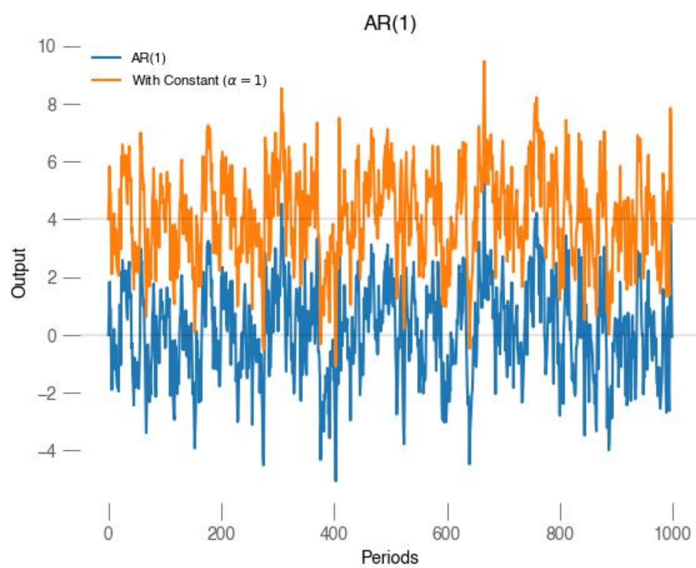
Importantly, neither of these two time series are stationary. A time series is **stationary** if:

1. $\mathbb{E}[y_t] = \mu$ for all t
2. $Var(y_t) < \infty$ for all t
3. $Cov(x_s, x_t) = Cov(x_{s+h}, x_{t+h})$

Let's see why these random walks fail to be stationary. Calculate the variance of the AR(1) process:

$$Var(y_t) = Var(\alpha + \rho y_{t-1} + \varepsilon_t)$$
$$Var(y_t) = \rho^2 Var(y_{t-1}) + \sigma^2$$
$$Var(y) = \frac{\sigma^2}{1-\rho^2}$$

So, obviously, if $\rho = 1$, the variance explodes to infinity. This proves that the random walks are non-stationary using contradiction.[1] Now let's look at the AR(1) process with $\rho = 0.75$:



Notice how both of these hover around their means (the light grey lines) and how neither seem to explode away from their means. These are both stationary series. The standard AR(1) has a mean of zero, so let's find the mean of the AR(1) with a constant:

$$\begin{aligned}\mathbb{E}[y_t] &= \mathbb{E}[\alpha + \rho y_{t-1} + \varepsilon_t]\\ \mu &= \alpha + \rho\mu\\ \mu &= \frac{\alpha}{1-\rho}\\ &= 4\end{aligned}$$

[1]Technically, ρ can be greater than one and the model be stationary. However, it is stationary in a future causal sense – that is, instead of y_{t-1} causing y_t, y_t causes y_{t-1}. Practically, we restrict $|\rho| < 1$.

which matches the mean on the AR(1) with constant time series.

We can also take expectations conditioning on time, using the following notation:

$$\mathbb{E}[x|\mathcal{F}_{t-1}] = \mathbb{E}_{t-1}[x]$$

where $\mathcal{F}$ denotes the information set up to and including time $t-1$. Typically, I prefer the notation on the right-hand side for its simplicity. Let's take the expectation of y_{t+1} given we know everything up to $t-1$:

$$\begin{aligned}
\mathbb{E}_{t-1}\left[y_{t+1}\right] &= \mathbb{E}_{t-1}\left[\alpha + \rho y_t + \varepsilon_t\right] \\
&= \alpha + \rho\mathbb{E}_{t-1}\left[\alpha + \rho y_{t-1} + \varepsilon_{t-1}\right] \\
&= \alpha + \rho\alpha + \rho^2 y_{t-1} + \rho\varepsilon_{t-1}
\end{aligned}$$

The last thing we will do for this section is recursively substitute in for y_{t-j}. Start with the standard AR(1) equation:

$$\begin{aligned}
y_t &= \alpha + \rho y_{t-1} + \varepsilon_t \\
&= \alpha + \rho(\alpha + \rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= (1+\rho)\alpha + \rho^2(\alpha + \rho y_{t-3} + \varepsilon_{t-2}) + \rho\varepsilon_{t-1} + \varepsilon_t \\
&= (1+\rho+\rho^2)\alpha + \rho^3 y_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
y_t &= \alpha\sum_{i=0}^{\infty}\rho^i + \sum_{i=0}^{\infty}\rho^i\varepsilon_{t-i}
\end{aligned}$$

Now, assume that the time series is stationary and use the sum of a geometric sequence:

$$y_t = \frac{\alpha}{1-\rho} + \sum_{i=0}^{\infty}\rho^i\varepsilon_{t-i}$$

Now plug in the mean of an AR(1):

$$y_t = \mu + \sum_{i=0}^{\infty}\rho^i\varepsilon_{t-i}$$

This form of an AR(1) is known as the **MA(∞)** form. Basically, the value of variable y_t is entirely determined by its mean and the decaying sum of all past shocks.

3.1 Brownian Motion

At some point in the past, you may have heard of Brownian motion. **Brownian motion**, denoted by $B(t)$ with $t \geq 0$, is a continuous time process with the following properties:

1. At $t = 0$, $B(0) = 0$
2. Displacement over interval (t_i, t_j), written as $B(t_i) - B(t_j)$, is normally distributed with mean zero and variance $\alpha(t_1 - t_0)$, where α is a scaling parameter
3. Displacements over non-overlapping intervals are independent
4. $B(t)$ is almost surely continuous (you will go over types of convergence with Drew)

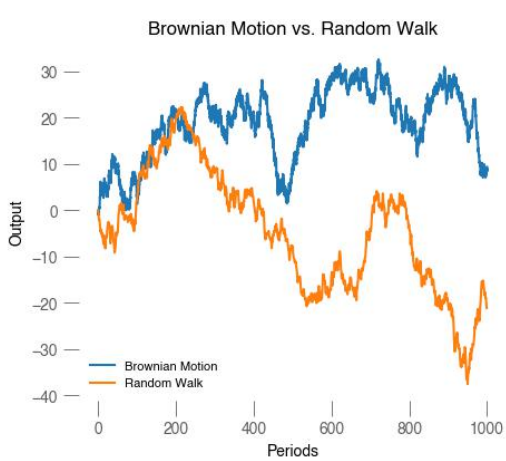
This definition can sound complicated. It turns out, however, that Brownian motion is just the continuous time limit of our discrete time random walk. How? Suppose the steps of the random walk happen at Δt intervals:

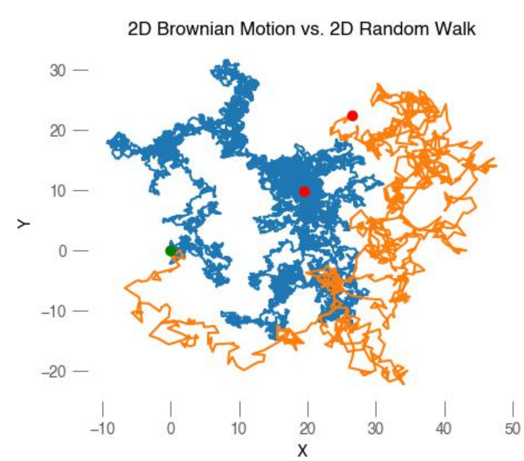
$$y_t = y(t) = y\left(\frac{t}{\Delta t}\right)$$

Now let that Δt go to zero. It turns out that:

$$B(t) = \sqrt{\Delta t}y(t)$$

I plot two examples of Brownian motion below:





Note that the behavior between Brownian motion and a random walk is essentially the same. The only difference is in the number of steps that each takes over the time horizon.