Intermediate Probability

Alex Houtz*

July 27, 2023

1 Multivariate Probability

1.1 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 1. 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:

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

*Math Camp Instructor | University of Notre Dame

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 2. 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 the same as always:

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

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:

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

So:

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

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.

1.2 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:

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

Conditional distributions leads to a very powerful theorem:

Theorem 1 (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 2 (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:

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

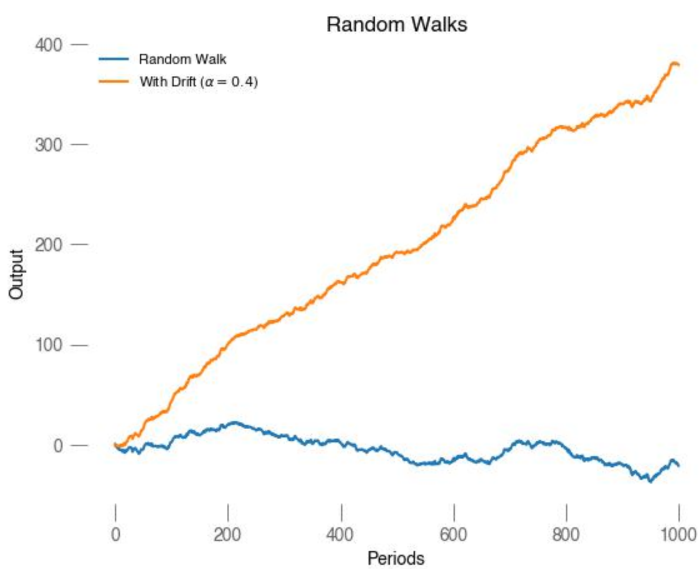
which is the law of total variance. ■

2 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 distributed *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 following picture:



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:

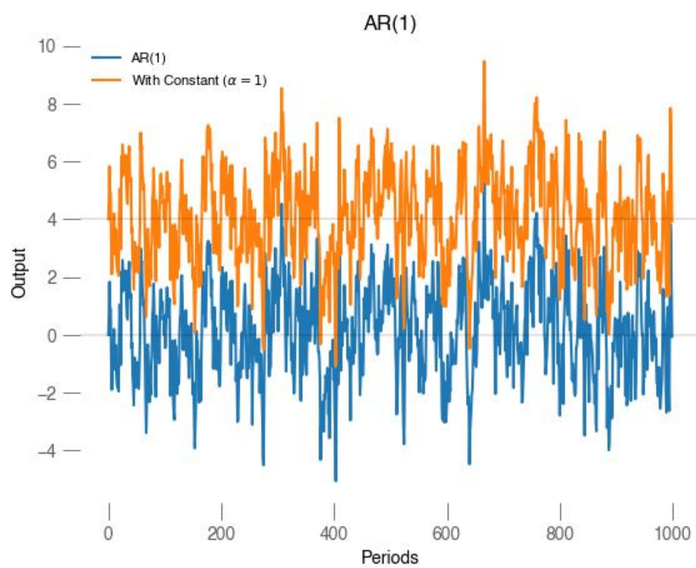
1. $\mathbb{E}[y_t] = \mu$ for all t
2. $Var(y_t) < \infty$ for all t
3. $Cov(x_s, x_t) = Cov(x_{s+h}, x_{t+h})$

Let's see why these random walks fail to be stationary. Calculate the variance of the

AR(1) process:

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

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.