No class or office hours on Tuesday, February 17.
Homework 1 due Friday, February 27.

Based on the iid properties of $Z^{(n)}$, if the time step $\Delta t$ is large enough, it suffices to complete the description of the probability model to give a probability distribution for a random kick $Z$; each $Z^{(n)} \sim Z$.

Now $Z$ has multiple components corresponding to the magnitude of the kick along the various coordinate directions; no a priori reason we should take these as independent (you're asked to think about whether the particular model on the homework has this property.)

So then how do we describe the random vector $Z$? A full description is the joint PDF $p_Z(z) = p_Z(z_1, z_2, \ldots, z_d)$ where $d$ is the number of spatial dimensions (typically $d = 2$ or $d = 3$).

Just as for single random variables, we can summarize basic properties of multiple random variables (i.e., the components of a random vector) by looking at their first and second moments.

The most fundamental summary statistic is still the mean:

$$E[Z] \equiv \mu_Z \equiv \langle Z \rangle = \begin{bmatrix} \mathbb{E}[Z_1] \\ \mathbb{E}[Z_2] \\ \vdots \\ \mathbb{E}[Z_d] \end{bmatrix}$$

So the mean of a random vector is just a vector whose components are the means of the components; there is no information involving relationship between the components.

However, the second moment has a nontrivial generalization from single random variables. The generalization of the variance is the covariance matrix:

$$\text{Cov}(Z, Y) = \mathbb{E}\left( (Z - \mu_Z) \otimes (Y - \mu_Y) \right)$$

where the tensor product is defined so that $A \otimes B$ is a matrix whose components are $A_i B_j$.

So the matrix $\text{Cov}(Z, Y)$ is a $d \times d$ matrix whose $(i, j)$ entry is:
\[ \text{Cov}(Z_i, Y_j) = \mathbb{E}\left( (Z_i - \mu_{Z_i})(Y_j - \mu_{Y_j}) \right) \]

This scalar covariance characterizes the relationship between the fluctuations of the two arguments about their means.

- covariance is zero if (but not only if) \( Z_i, Y_j \) are independent.

\[
\begin{align*}
\mathbb{E}\left( (Z_i - \mu_{Z_i})(Y_j - \mu_{Y_j}) \right) & = \mathbb{E}\left( Z_i - \mu_{Z_i} \right) \mathbb{E}\left( Y_j - \mu_{Y_j} \right) \\
& = \left( \mathbb{E}\left( Z_i - \mu_{Z_i} \right) \right) \left( \mathbb{E}\left( Y_j - \mu_{Y_j} \right) \right) \\
& = \left( \mathbb{E}\left( Z_i - \mu_{Z_i} \right) \right) \left( \mathbb{E}\left( Y_j - \mu_{Y_j} \right) \right) \\
& = \left( \mu_{Z_i} - \mu_{Z_i} \right) \left( \mu_{Y_j} - \mu_{Y_j} \right) \\
& = (0)(0) = 0
\end{align*}
\]

- covariance is positive when the fluctuations in the two arguments tend to be in the same direction
- covariance is negative when the fluctuations in the two arguments tend to be in the opposite direction

Can't tell the importance of the relationship between random variables by just looking at the covariance; need to look at the Pearson correlation coefficient:

\[ \rho(Z_i, Y_j) = \frac{\text{Cov}(Z_i, Y_j)}{\sigma_{Z_i}\sigma_{Y_j}} \]

This is like a normalized (nondimensionalized) covariance:

\[ -1 \leq \rho(Z_i, Y_j) \leq 1 \]

Values near \( \pm 1 \) represent strong correlations (\( \pm 1 \) means essentially perfect correlation). It's essentially the slope of the best fit line to the scatterplot of data for the two random variables (after normalization).
(linear regression). Probably a better way to think of the correlation coefficient is that it describes how well the scatterplot of the data can be fit by a linear regression.

The covariance matrix is simply a matrix of covariances between each component of $Z$ and each component of $Y$. One could similarly define a correlation matrix whose entries are $p(Z_i, Y_j)$.

Of particular interest is the covariance matrix of a random vector with itself: $\text{Cov}(Z, Z)$.

This is matrix whose $(i, j)$ entry is $\text{Cov}(Z_i, Z_j)$. Along the diagonal:

$$\text{Cov}(Z_i, Z_i) = \text{Var}(Z_i)$$

so the covariance matrix looks like:

$$\text{Cov}(Z) = \begin{pmatrix}
\text{Var}(Z_i) & \text{Cov}(Z_i, Z_j) & \cdots \\
\text{Cov}(Z_j, Z_i) & \text{Var}(Z_j) & \cdots \\
\vdots & \vdots & \ddots & \cdots \\
\text{Cov}(Z_d, Z_i) & \cdots & \cdots & \text{Var}(Z_d)
\end{pmatrix}$$

Properties of covariance matrices:

They are bilinear operators. What does this mean?

It means that if we have random variables $\{Y^i\}_{i=0}^m$ and constants $\{c^{(i)}, d^{(i)}\}_{i=0}^m$ and constant vectors $a, b$, where all these constants are deterministic.

Then:

$$\text{Cov}\left(\sum_{i=1}^m c^{(i)} Y^i + a, \sum_{j=1}^m d^{(j)} Z^j + b\right)$$

$$= \sum_{i=1}^m \sum_{j=1}^m c^{(i)} d^{(j)} \text{Cov}(Y^i, Z^j)$$
Then:

\[
\sum_{i=1}^{m} c(i) \sum_{j=1}^{m} d(j) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i)
\]

(note that the addition of deterministic constants to a random variable does not change its covariance with other random variables; this is an extra condition separate from the general technical definition of bilinearity)

Linearity of expectation:

\[
E \left( \sum_{i=1}^{m} c(i) \sum_{j=1}^{m} d(j) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i) \rightarrow \sum_{i=1}^{m} d(j) \sum_{j=1}^{m} c(i) \right)
\]

For covariance matrices \(Cov(Z, Z)\):

- symmetric
- positive semidefinite matrix

Why positive semidefinite? Use bilinearity of covariance, choose arbitrary deterministic vector \(\xi \in \mathbb{R}^d\):

\[
\xi \cdot Cov(Z, Z) \cdot \xi = Cov(\xi^T \xi, Z) \leq 0
\]

We talked about three standard one-dimensional random variable models for continuous random variables. Interestingly, it is somewhat difficult to have flexible models for multidimensional random vectors with arbitrary specification of the covariance of the components of the random vector. Normal (Gaussian) distributions have the most natural multidimensional
A Gaussian random vector $Y$ (viewed as a vector of random variables in each component) has a joint pdf:

$$p_Y(y) = \frac{1}{m} \frac{1}{(2\pi)^{\frac{m}{2}} (\det C)^{\frac{1}{2}}} \exp \left( -\frac{1}{2} (y - \mu) \cdot C^{-1} \cdot (y - \mu) \right)$$

where the parameters are:

- $\mu$: the mean of the random vector $\mu = EY$
- $C$: the covariance matrix of the random vector: $C = Cov(Y,Y)$

and $m$ is the number of components/dimensions of $Y$.

A multidimensional Gaussian/normal distribution is well-specified for any choice of $\mu \in \mathbb{R}^m$ and positive definite, symmetric $C$.

This ability does not extend well to other families of probability distributions; this is why one often uses a "Gaussian copula" to remap other probability distributions for multiple random variables to a Gaussian distribution (but there is some cheating/approximation involved).

**Conditional Probability**

A more detailed way (relative to covariance matrix) to discuss the relationship between two random variables $Y$ and $Z$ is by asking how knowing the realization of one affects the probability distribution for the other.

More primitively, we say the conditional probability of an event $A$ given an event $C$ is:

$$P(A|C) = \frac{P(A \cap C)}{P(C)}$$

We can apply this to do conditional calculations with random variables, first when they are discrete:

$$P(\tilde{Z} = \tilde{z} | \tilde{Y} = \tilde{y}) = \frac{P(\tilde{Z} = \tilde{z}, \tilde{Y} = \tilde{y})}{P(\tilde{Y} = \tilde{y})}$$
Generalizing conditional probability to continuous random variables is a technical pain to do rigorously. Fortunately the result is simple and analogous to the discrete case. (One just has to be a bit careful in using conditioning on continuous random variables; see for example the Borel paradox.)

One simply defines the conditional probability density for a random variable $Z$ given the value of a random variable $Y$ by dividing the joint probability density by the marginal probability density for the given random variable:

$$p_{Z|Y}(z|y) = \frac{p_{Z,Y}(z,y)}{p_Y(y)}$$

This object is how you adjust the probability density for $Z$ when you are given information that $Y = y$. Actually to be extra careful, this conditional probability density really is assuming that $Y$ is close to $y$, not exactly equal to it.