Probability and Averaging

Now that we have developed the necessary tools to describe random variables, random processes, and random fields, it is useful to return to our starting point to relate concepts developed here to measurable quantities such as ensemble, time, and spatial averages.

Let $u(t)$ be the specified component of velocity at a specified position for time t as measured by an experiment. For each experiment we conduct, we observe a different value for $u(t)$.

Let $u^{(i)}(t)$ be the i^{th} observation or realization of $u(t)$. Like $u(t)$, $u^{(i)}(t)$ is a random variable, and we may assume each realization is independent and identically distributed.

We define the ensemble average as:

Ensemble Average: $$\langle u(t) \rangle_N \equiv \frac{1}{N} \sum_{n=1}^{N} u^{(n)}(t)$$

which itself is a random variable, and if each realization is distributed like $u(t)$, then:

$$\langle \langle u(t) \rangle_N \rangle = \langle u(t) \rangle$$

$$\mathrm{var}\left(\langle u(t) \rangle_N\right) = \frac{1}{N} \mathrm{var}\left(u(t)\right)$$

Consequently, the ensemble average converges almost surely to the mean as $N \to \infty$ in the sense that it converges with 100% probability. That is:

$$P\left(\lim_{N \to \infty} \langle u(t) \rangle_N = \langle u(t) \rangle \right) = 1$$

where we note the limit itself is a random variable!

The above discussion is incredibly powerful because it says we can replace something we cannot easily measure (the mean) with something we can (the ensemble average).

The central limit theorem expands on this, stating that:

$$\lim_{N \to \infty} P\left\langle \left| \frac{[\langle u(t) \rangle_N - \langle u(t) \rangle]}{\underbrace{\sigma_u}_{\mathrm{var}^{1/2}(u(t))}} \right| < \frac{\varepsilon}{N^{1/2}} \right\rangle = \frac{1}{2\pi} \int_{-\infty}^{\varepsilon} e^{-z^2/2} \, dz$$

for every fixed $\varepsilon > 0$.

The above estimate demonstrates that ensemble averages converge to the mean with a convergence rate of $N^{-1/2}$. Monte Carlo methods are predicated on this fact.

Note: There is one caveat in the above analysis. The variance must be finite!

Of course, the ensemble average is just one of many potential averages. For statistically stationary flows, we may consider the time average over a time interval T:

Time-Average: $$\langle u(t) \rangle_T = \frac{1}{T} \int_{t}^{t+T} u(t') \, dt'$$

Like the ensemble average, the time-average is a random variable with mean:

$$\langle\langle u(t)\rangle_T\rangle = \langle u(t)\rangle$$

However, we do not immediately know if the time average converages almost surely to the mean. A random process whose time average does converge almost surely to the mean is said to be mean-ergodic.

The ergodicity assumption underlies almost all measurements of turbulent statistics and much of the theory underlying Reynolds Averaged Navier-Stokes models. Nonetheless, there are no proofs that statistically stationary turbulent flows are ergodic. The moral of the story is to understand the assumptions behind turbulence modeling and their limitations!

IF a random process $u(t)$ ~~has~~ is statistically stationary and has finite integral time-scale:

$$\bar{\tau} = \int_0^\infty \rho(s)\,ds \qquad \rho(s) = \frac{\langle u'(t)\,u'(t+s)\rangle}{\langle (u'(t))^2\rangle}$$

then the process is necessarily mean-ergodic! To see this, we formally write:

$$\lim_{T\to\infty} \mathrm{var}\left(\langle u(t)\rangle_T\right) = \lim_{T\to\infty} \left\langle \left(\langle u(t)\rangle_T'\right)^2\right\rangle$$

where:

$$\langle u(t)\rangle_T' = \langle u(t)\rangle_T - \langle u(t)\rangle$$

$$= \frac{1}{T}\int_{t}^{T} u(t')\,dt' - \frac{1}{T}\int_0^T \langle u(t')\rangle\,dt'$$

$u(t') - u'(t')$

Assume $t=0$ to simplify our proceeding analysis!

$$= \frac{1}{T}\int_0^T u'(t')\,dt'$$

Consequently, the fluctuation of the time-average is the time-average of the fluctuation. Thus:

$$\lim_{T\to\infty} \mathrm{var}\left(\langle u(t)\rangle_T\right) = \lim_{T\to\infty} \left\langle \frac{1}{T^2}\int_0^T\int_0^T u'(t_1)\,u'(t_2)\,dt_1\,dt_2\right\rangle$$

$$= \lim_{T\to\infty} \frac{1}{T^2}\int_0^T\int_0^T \langle u'(t_1)\,u'(t_2)\rangle\,dt_1\,dt_2$$

$$= \lim_{T\to\infty} \frac{1}{T^2}\mathrm{var}(u(t)) \int_0^T\int_0^T \rho(t_2-t_1)\,dt_1\,dt_2$$

Exploits, among other things, Fubini's!

$$= \lim_{T\to\infty} \frac{1}{T^2}\mathrm{var}(u(t)) \int_0^T\int_{-t_1}^{T-t_1} \rho(s)\,ds\,dt_1$$

The tricky step!

$$\leq \lim_{T\to\infty} \frac{\mathrm{var}(u(t))}{T^2}\cdot\int_0^T 2\bar{\tau}\,dt_1$$

$$= \lim_{T\to\infty} \frac{2\,\mathrm{var}(u(t))\,\bar{\tau}}{T} = 0$$

Above, note that for the variance of the time average to be small, the time interval T had to be large compared with the integral time-scale $\bar{\tau}$!

Moreover, for a process with finite integral time-scale and variance, it can be shown that the time average converges like $T^{-1/2}$, $\bar{\tau}^{1/2}$, and $\text{var}^{1/2}(u(t))$!

So what does all of the above mean? If the integral time-scale is finite, then there is a finite time at which a sample path forgets its initial condition. As a result, the time-averaged statistics past this point are invariant with respect to the initial condition and hence we obtain an ergodic process.

On the other hand, if the time-scale is infinite, a sample path may never forget its initial condition and the time-averaged statistics will necessarily depend on the initial condition. In this situation, we may not substitute the time average for the mean. !!!

Finally, for homogeneous turbulent flow fields, we may define the spatial average of $\vec{u}(\vec{x},t)$ as:

$$\langle \vec{u}(\vec{x},t) \rangle \equiv \frac{1}{L^3} \int_0^L \int_0^L \int_0^L \vec{u}(\vec{x},t)\, dx_1\, dx_2\, dx_3$$

and note that L defines the length of the sides of the domain, a cube.