Plan

0. Brief Review and Questions
1. Optimal Predictors and Conditional Expected Values
2. Stochastic Processes Defined
3. Fun with White Noise Part I
   - The White Noise Process
   - The Random Walk
   - The Poisson Process

Next Time: More Fun with White Noise

Reading: G&S 3.9, 3.10, 5.1, 5.2 (optional), 5.3
Homework 2 due next Tuesday

Review Definition 1. Let Y be a scalar-valued random variable and X be an arbitrary (possibly vector-valued) random variable.

- A *predictor* of Y given X is a function that maps each possible value of X (that is, each value in the range of X) to a real number. This number represents our guess of the value of Y if the corresponding value of X is observed.
- A *prediction* of Y given X is the random variable that represents the guess that will be made, using a particular predictor, when X is eventually observed.
- The optimal (mean square) predictor of Y given a (possibly vector-valued) random variable X, $\mathsf{pred}_{Y|X}$, is the function $g \in \mathcal{G}$ that minimizes $\mathsf{E}(Y - g(X))^2$. The optimal (mean square) prediction is the *random variable* $g(X)$ for that same g (that is $\mathsf{pred}_{Y|X}(X)$).

Example 2. Optimal predictors:

1. If X is a constant, $\mathsf{pred}_{Y|X}(u) = \mathsf{E}Y$.
2. If X is an indicator,
$$\mathsf{pred}_{Y|X}(u) = \begin{cases} \dfrac{\mathsf{E}Y(1-X)}{\mathsf{E}(1-X)} & \text{if } u = 0 \\ \dfrac{\mathsf{E}YX}{\mathsf{E}X} & \text{if } u = 1. \end{cases} \tag{1}$$
3. If X is a discrete random variable with PMF p_X,
$$\mathsf{pred}_{Y|X}(u) = \frac{\mathsf{E}Y\,1\{X = u\}}{\mathsf{p}_X(u)}. \tag{2}$$
4. If X is a continuous random variable with PDF f_X, we can write (only somewhat fancifully):
$$\mathsf{pred}_{Y|X}(u) = \frac{\mathsf{E}Y\,1\{X \text{ near } u\}}{\mathsf{P}\{X \text{ near } u\}}, \tag{3}$$
   where the $\{X \text{ near } u\}$ denotes (again somewhat fancifully) the event $\{X \in [u, u + du)\}$.

Heuristic Definition 3. This leads to definitions of several useful and common quantities:

1. Conditional probabilities

$$\mathsf{P}(B \mid A) = \frac{\mathsf{E}1_B 1_A}{\mathsf{E}1_A} = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}. \tag{4}$$

2. The *local conditional expectation* operator:

$$\mathsf{E}(Y \mid X \text{ near } u) = \mathsf{pred}_{Y|X}(u) \tag{5}$$

that holds for real- and vector-valued random variables. The operator $\mathsf{E}(\cdot \mid X \text{ near } u)$ satisfies all the basic expected value rules.

3. Conditional distributions

$$\mathsf{D}_{Y|X}(h \mid u) = \mathsf{E}(h(Y) \mid X \text{ near } u). \tag{6}$$

4. Conditional distributions, for example PMFs and PDFs. The conditional PMF $\mathsf{p}_{Y|X}$ or PDF $\mathsf{f}_f Y \mid X$ is defined such that

$$\mathsf{E}(h(Y) \mid X = u) = \sum_v h(v)\mathsf{p}_{Y|X}(v \mid u)$$

$$\mathsf{E}(h(Y) \mid X \text{ near } u) = \int h(v)\mathsf{f}_{Y|X}(v \mid u)\, dv$$

A useful mnemonic device is "joint equals conditional times marginal."

Heuristic Definition 4. The *conditional expectation* of Y given X is a random variable $\mathsf{E}(Y \mid X)$ given by

$$\mathsf{E}(Y \mid X) = \mathsf{pred}_{Y|X}(X). \tag{7}$$

This is the optimal prediction of Y given the observed value of X.

Formal Definition (Part I) 5. Let $(\Omega, \mathcal{F}, \mathsf{E})$ be a probability space and let $\mathcal{G} \subset \mathcal{F}$ be a σ-field. If Y is a real-valued random variable such that $\mathsf{E}Y$ exists, then there exists a random variable , denoted by $\mathsf{E}(Y \mid \mathcal{G})$, with the following properties:

1. $\mathsf{E}(Y \mid \mathcal{G})$ is $\mathcal{G}$-measurable (meaning $\left\{\mathsf{E}(Y \mid \mathcal{G}) \in A\right\} \in \mathcal{G}$ for every Borel or null set A).
2. $\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G})\right)$ exists
3. For every $G \in \mathcal{G}$,

$$\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G})1_G\right) = \mathsf{E}\left(Y 1_G\right). \tag{8}$$

This extends naturally to vector-valued variables

(Rigor alert: the random variable $\mathsf{E}(Y \mid \mathcal{G})$ is *not* unique, but it is unique up to events of probability zero. Moreover, we can find one version of this random variable with all the regularity properties we would expect, just i's and t's, folks.)

Claim 6. If $G\colon (\Omega, \mathcal{F}) \to (\mathcal{R}, \mathcal{B})$ is $\mathcal{G}$-measureable (meaning that $G^{-1}(B) \in \mathcal{G}$ for $B \in \mathcal{B}$) and if $\mathsf{E}G$ and $\mathsf{E}YG$ exist, then

$$\mathsf{E}G(Y - \mathsf{E}(Y \mid \mathcal{G})) = 0. \tag{9}$$

Formal Definition (Part II) 7. Assume the conditions of the previous definition. Let X be a random variable (real or vector-valued) and define $\mathcal{G} = \{X^{-1}(A)\}$ for sets in the corresponding σ-field in the range of X. Then $\mathcal{G}$ is a σ-field and we define

$$\mathsf{E}(Y \mid X) = \mathsf{E}(Y \mid \mathcal{G}), \tag{10}$$

where we can assume we've picked a "nice" version. The defining property of the conditional expectation corresponds to the following useful identity for any (measurable) function g:

$$\mathsf{E}\left(\mathsf{E}(Y \mid X)g(X)\right) = \mathsf{E}\left(Yg(X)\right), \tag{11}$$

or more compellingly

$$\mathsf{E}g(X)(Y - \mathsf{E}(Y \mid X)) = 0. \tag{12}$$

Claim 8. The formal and heuristic definitions of $\mathsf{E}(Y \mid X)$ coincide for all practical purposes when $\mathsf{E}Y^2 < \infty$.

Claim 9. Both $\mathsf{E}(\cdot \mid X)$ and $\mathsf{E}(\cdot \mid \mathcal{G})$ satisfy the basic expected value rules.

Identity 10. The Enhanced Scaling Rule
If $\mathcal{G}$ is a σ-field of events and G is a $\mathcal{G}$-measurable random variable then

$$\mathsf{E}(GY \mid \mathcal{G}) = G\,\mathsf{E}(Y \mid \mathcal{G}) \tag{13}$$

Or, more simply, in terms of random variables

$$\mathsf{E}(g(X)Y \mid X) = g(X)\,\mathsf{E}(Y \mid X). \tag{14}$$

In both cases, I'm assuming that the relevant expected values exist.

Identity 11. The Mighty Conditioning Identity (Basic form)

$$\mathsf{E}\left(\mathsf{E}(Y \mid X)\right) = \mathsf{E}Y \tag{15}$$

This follows immediately from the formal definition above and comes out from the optimal predictor definitions as well.

Identity 12. The Mighty Conditioning Identity (General Form)
Suppose that $\mathcal{G}_1 \subset \mathcal{G}_2$ are σ-fields of events. Then, the following is true:

$$\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G}_1) \mid \mathcal{G}_2\right) = \mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G}_2) \mid \mathcal{G}_1\right) = \mathsf{E}(Y \mid \mathcal{G}_1). \tag{16}$$

Example 13. A random number of random variables.

Another Definition 14. Independence. Two random variables X and Y are independent iff

$$\mathsf{E}(h(Y) \mid X) = \mathsf{E}h(Y) \tag{17}$$

for all (measurable) functions h defined on the range of Y.

Definition 15. Stochastic Process I – Collection of Random Variables

Let $\mathcal{T}$ be an index set and $(\mathcal{X}, \mathcal{G})$ be a measurable space. A *stochastic process* is a collection of $\mathcal{X}$-valued random variables $(X_t)_{t \in \mathcal{T}}$.

Note 16. Unless otherwise indicated, we will assume that we are working in a probability space $(\Omega, \mathcal{F}, \mathsf{E})$ without mentioning it explicitly.

Examples 17. (a) Single Random Variable, (b) Random Vector, (c) IID Sequence, (d) General Sequence, (e) One-Dimensional Random Field, (f) Two-Dimonsional Random Field.

Definition 18. Stochastic Process II – Random Functions

Let $\mathcal{T}$ be an index set and $(\mathcal{X}, \mathcal{G})$ be a measurable space. A *stochastic process* X is random function defined on $\mathcal{T}$ and taking values in $\mathcal{X}$; that is, $X\colon \Omega \to \mathcal{X}^T$. The realized functions $X(\omega)$ are called *sample paths* of the process.

Notation 19. We will write the values X_t and $X(t)$ interchangeably as is most convenient. The realized values are $X_t(\omega)$ or $X(t, \omega)$ equivalently.

Rigor Alert 20. In order to be precise, X needs to be defined in some measurable space $(\mathcal{X}^T, \mathcal{H})$ for some σ-field $\mathcal{H}$. With proper attention to detail, we can construct the σ-field, but we will discuss that only as needed in this course.

Process 21. Let $(X_n)_{n \in \mathbb{Z}_\oplus}$ be a sequence of independent, identically distributed, mean zero random variables. We will call this a *discrete-time white noise process.*

Examples 22. (a) Bernoulli process, (b) Radamacher process, (c) Gaussian white noise, (d) Exponential waiting times.

Process 23. Let (X_n) be a white noise process and S_0 be an arbitrary random variable. Then, $(S_n)_{n \in \mathbb{Z}_\oplus}$, defined by

$$S_n = S_0 + \sum_{i=1}^{n} X_i, \tag{18}$$

is called a *random walk.*

Examples 24.

1. Coin Flips
2. Simple Random Walk
3. First return to zero.
4. Hitting Time
5. Gambler's Ruin (several ways)
6. Poisson Process.