Plan

0. Brief Review and Questions
1. One More Generating Function Example
2. Prediction and Expected Value
3. Homework Questions

Next Time: Fun with White Noise

Reading: G&S 4.1-4.7

Comment 1. Suppose $(\Omega, \mathcal{F}, \mathsf{E})$ is a probability space and $A \in \mathcal{F}$ has $\mathsf{P}(A) \equiv \mathsf{E}1_A = 0$. A is called a *null* set. It turns out that we can without any problem extend $\mathcal{F}$ to include all $B \subset A$ which are also necessarily null sets. We will assume henceforth, unless otherwise noted, that any σ-field we use has been extended in this way.

Review Definition 2. If X is a random variable, the *distribution* of X is the operator D_X that operates on (measurable) functions $g: \mathcal{X} \to \mathbb{R}^k$, for any $k \geq 1$ and $\mathcal{X}$ containing range(X), by

$$\mathsf{D}_X g = \mathsf{E} g(X). \tag{1}$$

Notice that the this definition of D_X transparently handles the case of vector-valued X. If $X = (X_1, \ldots, X_k)$ for some integer $k > 1$, then $\mathsf{D}_X g = \mathsf{E} g(X_1, \ldots, X_n)$. Given a collection of random variables $Z_1, \ldots, Z_m$, the *joint* distribution $\mathsf{D}_{Z_1,\ldots,Z_m}$ just the distribution D_Z where $Z = (Z_1, \ldots, Z_m)$.

Review Definition 3. Representations of Probability Distributions

Representation	Notation	How to get from D_X
Measure	μ_X	$\mu_X(\mathcal{A}) = \mathsf{P}\{X \in \mathcal{A}\} = \mathsf{D}_X 1_{\mathcal{A}}$
PMF	p_X	$\mathsf{p}_X(u) = \mathsf{P}\{X = u\} = D_X 1_{\{u\}}$
PDF	f_X	$\mathsf{f}_X(u) = \lim_{\Delta \to 0} \frac{1}{\Delta} \mathsf{D}_X 1_{[u,u+\Delta)}$
CDF	F_X	$\mathsf{F}_X(u) = \mathsf{P}\{X \leq u\} = \mathsf{D}_X 1_{(-\infty,u]}$
SDF	S_X	$\mathsf{S}_X(u) = \mathsf{P}\{X > u\} = \mathsf{D}_X 1_{]u,\infty[}$
PGF	G_X	$\mathsf{G}_X(z) = \mathsf{E} z^X = \mathsf{D}_X g_z$, where $g_z(u) = z^u$
MGF	M_X	$\mathsf{M}_X(s) = \mathsf{E} e^{-sX} = \mathsf{D}_X h_s$, where $h_s(u) = e^{su}$
CGF	C_X	$\mathsf{C}_X(t) = \mathsf{E} e^{itX} = \mathsf{D}_X r_t$, where $r_t(u) = e^{itu}$

Review (Working) Definition 4. We will say that random variables $X_1, \ldots, X_n$ are *independent* if for any (measurable) real-valued functions $g_1, \ldots, g_n$ defined on the respective ranges of the X_is,

$$\mathsf{E} \prod_{i=1}^{n} g_i(X_i) = \prod_{i=1}^{n} \mathsf{E} g_i(X_i). \tag{2}$$

Examples 5. Two generating function examples:

1. Pentagon walk (questions)
2. Double heads

Definition 6. 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 7. Optimal predictors:

1. If X is a constant, $\mathsf{pred}_{Y|X}(u) = \mathsf{E}Y$.
2. If X is an indicator,
$$\mathsf{pred}_{Y|X}(u) = \begin{cases} \dfrac{\mathsf{E}Y(1-X)}{\mathsf{E}(1-X)} & \text{if } u = 0 \\ \dfrac{\mathsf{E}YX}{\mathsf{E}X} & \text{if } u = 1. \end{cases} \tag{3}$$
3. If X is a discrete random variable with PMF p_X,
$$\mathsf{pred}_{Y|X}(u) = \frac{\mathsf{E}Y\,1\{X = u\}}{\mathsf{p}_X(u)}. \tag{4}$$
4. If X is a continuous random variable with PDF f_X, we can write (only somewhat fancifully):
$$\mathsf{pred}_{Y|X}(u) = \frac{\mathsf{E}Y\,1\{X \text{ near } u\}}{\mathsf{P}\{X \text{ near } u\}}, \tag{5}$$
where the $\{X \text{ near } u\}$ denotes (again somewhat fancifully) the event $\{X \in [u, u + du)\}$.

Heuristic Definition 8. This leads to definitions of several useful and common quantities:

1. Conditional probabilities
$$\mathsf{P}(B \mid A) = \frac{\mathsf{E}1_B 1_A}{\mathsf{E}1_A} = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(A)}. \tag{6}$$
2. The *local conditional expectation* operator:
$$\mathsf{E}(Y \mid X \text{ near } u) = \mathsf{pred}_{Y|X}(u) \tag{7}$$
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
$$\mathsf{D}_{Y|X}(h \mid u) = \mathsf{E}(h(Y) \mid X \text{ near } u). \tag{8}$$
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
$$\mathsf{E}(h(Y) \mid X = u) = \sum_v h(v)\mathsf{p}_{Y|X}(v \mid u)$$
$$\mathsf{E}(h(Y) \mid X \text{ near } u) = \int h(v)\mathsf{f}_{Y|X}(v \mid u)\,dv$$

Heuristic Definition 9. The *conditional expectation* of Y given X is a random variable $\mathsf{E}(Y \mid X)$ given by
$$\mathsf{E}(Y \mid X) = \mathsf{pred}_{Y|X}(X). \tag{9}$$
This is the optimal prediction of Y given the observed value of X.

Formal Definition (Part I) 10. 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. $\left\{\mathsf{E}(Y \mid \mathcal{G}) \in A\right\} \in \mathcal{G}$ for every Borel or null set A.
2. $\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G})\right)$ exists
3. For every $G \in \mathcal{G}$,
$$\mathsf{E}\left(\mathsf{E}(Y \mid \mathcal{G})1_G\right) = \mathsf{E}\left(Y 1_G\right). \tag{10}$$

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 11. If $G\colon(\Omega, \mathcal{F}) \to (\mathcal{R}, \mathcal{B})$ is $\mathcal{G}$-measureable (meaning that $G^{-1}(B) \in \mathcal{G}$) and if $\mathsf{E}G$ and $\mathsf{E}YG$ exist, then
$$\mathsf{E}G(Y - \mathsf{E}(Y \mid \mathcal{G})) = 0. \tag{11}$$

Formal Definition (Part II) 12. 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
$$\mathsf{E}(Y \mid X) = \mathsf{E}(Y \mid \mathcal{G}), \tag{12}$$
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:
$$\mathsf{E}\left(\mathsf{E}(Y \mid X)g(X)\right) = \mathsf{E}\left(Y g(X)\right), \tag{13}$$
or more compellingly
$$\mathsf{E}g(X)(Y - \mathsf{E}(Y \mid X)) = 0. \tag{14}$$

Claim 13. The formal and heuristic definitions of $\mathsf{E}(Y \mid X)$ coincide for all practical purposes when $\mathsf{E}Y^2 < \infty$.

Claim 14. Both $\mathsf{E}(\cdot \mid X)$ and $\mathsf{E}(\cdot \mid \mathcal{G})$ satisfy the basic expected value rules.

Identity 15. The Mighty Conditioning Identity
$$\mathsf{E}\left(\mathsf{E}(Y \mid X)\right) = \mathsf{E}Y \tag{15}$$
This follows immediately from the formal definition above and comes out from the optimal predictor definitions as well.

Example 16. A random number of random variables.

Another Definition 17. Independence. Two random variables X and Y are independent iff
$$\mathsf{E}(h(Y) \mid X) = \mathsf{E}h(Y) \tag{16}$$
for all (measurable) functions h on the suitable domain.