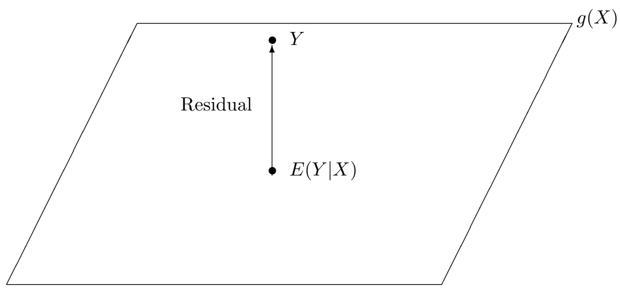
Stat 150 Stochastic Processes **Spring 2009**

Lecture 2: Conditional Expectation

Lecturer: Jim Pitman

- **Some useful facts** (assume all random variables here have finite mean square):
  • $\mathbb{E}(Yg(X)|X) = g(X)\mathbb{E}(Y|X)$
  • $Y - \mathbb{E}(Y|X)$ is orthogonal to $\mathbb{E}(Y|X)$, and orthogonal also to $g(X)$ for every measurable function g.
  • Since $\mathbb{E}(Y|X)$ is a measurable function of X, this characterizes $\mathbb{E}(Y|X)$ as the *orthogonal projection* of Y onto the linear space of all square-integrable random variables of the form $g(X)$ for some measurable function g.
  • Put another way $g(X) = \mathbb{E}(Y|X)$ minimizes the mean square prediction error $\mathbb{E}[(Y - g(X))^2]$ over all measurable functions g.



These facts can all be checked by computations as follows: Check orthogonality:

$$\begin{aligned}
\mathbb{E}[(Y - \mathbb{E}(Y|X))g(X)] &= \mathbb{E}(g(X)Y - g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(g(X)Y) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(\mathbb{E}(g(X)Y|X)) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= \mathbb{E}(g(X)\mathbb{E}(Y|X)) - \mathbb{E}(g(X)\mathbb{E}(Y|X)) \\
&= 0
\end{aligned}$$

- Recall: $\mathbf{V}ar(Y) = \mathbb{E}(Y - \mathbb{E}(Y))^2$ and $\mathbf{V}ar(Y|X) = \mathbb{E}([Y - \mathbb{E}(Y|X)]^2|X)$.
  Claim: $\mathbf{V}ar(Y) = \mathbf{V}ar(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbf{V}ar(Y|X))$

Proof:

$$
\begin{aligned}
Y &= \mathbb{E}(Y|X) + Y - \mathbb{E}(Y|X) \\
\mathbb{E}(Y^2) &= \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}([Y - \mathbb{E}(Y|X)]^2) + 0 \\
&= \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}(\mathbf{V}ar(Y|X)) \\
\mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 &= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(Y))^2 + \mathbb{E}(\mathbf{V}ar(Y|X)) \\
&= \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(\mathbb{E}(Y|X)))^2 + \mathbb{E}(\mathbf{V}ar(Y|X))
\end{aligned}
$$

$$\Longrightarrow \mathbf{V}ar(Y) = \mathbf{V}ar(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbf{V}ar(Y|X))$$

- Exercise P. 84, 4.3

  T is uniform on [0,1]. Given T, U is uniform on $[0, T]$. What is $\mathbb{P}(U \geq 1/2)$?

$$
\begin{aligned}
\mathbb{P}(U \geq 1/2) &= \mathbb{E}(\mathbb{E}[\mathbf{1}(U \geq 1/2)|T]) \\
&= \mathbb{E}[\mathbb{P}(U \geq 1/2|T)] \\
&= \mathbb{E}\left[\frac{T - 1/2}{T}\mathbf{1}(T \geq 1/2)\right] \\
&= \int_{1/2}^{1} \frac{t - 1/2}{t} dt
\end{aligned}
$$

- **Random Sums:** Random time T. $S_n = X_1 + \cdots + X_n$. Wants a formula for $\mathbb{E}(S_T)$ which allows that T might not be independent of $X_1, X_2, \ldots$.

  Condition: For all $n = 1, 2, \ldots$ the event $(T = n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T \leq n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T > n)$ is determined by $X_1, X_2, \ldots, X_n$.
  Equivalently: $(T \geq n)$ is determined by $X_1, X_2, \ldots, X_{n-1}$.
  Call such a T a *stopping time* relative to the sequence $X_1, X_2, \ldots$
  Example: The first n (if any) such that $S_n \leq 0$ or $S_n \geq b$. Then $(T = n) = (S_1 \in (0, b), S_2 \in (0, b), \ldots, S_{n-1} \in (0, b), S_n \notin (0, b))$ is a function of $S_1, \ldots, S_n$.

- **Wald's identity:** If T is a stopping time relative to $X_1, X_2, \ldots$, which are i.i.d. and $S_n := X_1 + \cdots + X_n$, then $\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1)$, provided $\mathbb{E}(T) < \infty$.

Sketch of proof:

$$\begin{aligned} S_T &= X_1 + \cdots + X_T \\ &= X_1\mathbf{1}(T \geq 1) + X_2\mathbf{1}(T \geq 2) + \cdots \end{aligned}$$

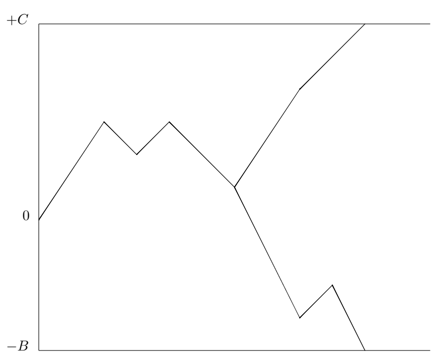
$$\begin{aligned} \mathbb{E}(S_T) &= \mathbb{E}(X_1\mathbf{1}(T \geq 1) + \mathbb{E}(X_2\mathbf{1}(T \geq 2)) + \cdots \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2)\mathbb{P}(T \geq 2) + \mathbb{E}(X_3)\mathbb{P}(T \geq 3) + \cdots \\ &= \mathbb{E}(X_1)\sum_{n=1}^{\infty}\mathbb{P}(T \geq n) \\ &= \mathbb{E}(X_1)\mathbb{E}(T) \end{aligned}$$

Key point is that for each n the event $(T \geq n)$ is determined by $X_1, X_2, \ldots, X_{n-1}$, hence is independent of X_n. This justifies the factorization

$$\mathbb{E}(X_n 1(T \geq n) = \mathbb{E}(X_n)\mathbb{E}(\mathbf{1}(T \geq n)) = \mathbb{E}(X_1)\mathbb{P}(T \geq n).$$

It is also necessary to justify the swap of $\mathbb{E}$ and Σ. This is where $\mathbb{E}(T) < \infty$ must be used in a more careful argument. Note that if $X_i \geq 0$ the swap is justified by monotone convergence.

- Example. Hitting probabilities for simple symmetric random walk



$S_n = X_1 + \cdots X_n$, $X_i \sim \pm 1$ with probability 1/2,1/2. T= first n s.t. $S_n = +C$ or $-B$.

It is easy to see that $\mathbb{E}(T) < \infty$. Just consider successive blocks if indices of length $B + C$. $\underbrace{\qquad\quad}_{B+C}\underbrace{\qquad\quad}_{B+C}\underbrace{\qquad\quad}_{B+C}\underbrace{\qquad\quad}_{\ldots}$ Wait until a block of length $B + C$ with $X_i = 1$ for all i in the block. Geometric distribution of this upper bound on $T \Longrightarrow \mathbb{E}(T) < \infty$.

Let $p+ = \mathbb{P}(S_T = +C)$ and $p_- := \mathbb{P}(S_T = -C)$. Then

$$\begin{aligned}
\mathbb{E}(S_T) &= \mathbb{E}(T)\mathbb{E}(X_1) = \mathbb{E}(T) \cdot 0 = 0 \\
0 &= p_+ C - p_- B \\
1 &= p_+ + p_- \\
\implies p_+ &= \frac{B}{B+C} \qquad p_- = \frac{C}{B+C}
\end{aligned}$$