• **Some useful facts** (assume all random variables here have finite mean square):
  - $E(Yg(X)|X) = g(X)E(Y|X)$
  - $Y - E(Y|X)$ is orthogonal to $E(Y|X)$, and orthogonal also to $g(X)$ for every measurable function $g$.
  - Since $E(Y|X)$ is a measurable function of $X$, this characterizes $E(Y|X)$ 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, $g(X) = E(Y|X)$ minimizes the mean square prediction error $E[(Y - g(X))^2]$ over all measurable functions $g$.

![Diagram](image)

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

$$E[(Y - E(Y|X))g(X)] = E(g(X)Y - g(X)E(Y|X))$$
$$= E(g(X)Y) - E(g(X)E(Y|X))$$
$$= E(E(g(X)Y|X)) - E(g(X)E(Y|X))$$
$$= E(g(X)E(Y|X)) - E(g(X)E(Y|X))$$
$$= 0$$

• Recall: $Var(Y) = E(Y - E(Y))^2$ and $Var(Y|X) = E((Y - E(Y|X))^2|X)$.

**Claim:** $Var(Y) = Var(E(Y|X)) + E(Var(Y|X))$
Proof:

\[ Y = \mathbb{E}(Y|X) + Y - \mathbb{E}(Y|X) \]
\[ \mathbb{E}(Y^2) = \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}([Y - \mathbb{E}(Y|X)]^2) + 0 \]
\[ = \mathbb{E}(\mathbb{E}(Y|X)^2) + \mathbb{E}(\text{Var}(Y|X)) \]
\[ \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(Y))^2 + \mathbb{E}(\text{Var}(Y|X)) \]
\[ = \mathbb{E}(\mathbb{E}(Y|X)^2) - (\mathbb{E}(\mathbb{E}(Y|X)))^2 + \mathbb{E}(\text{Var}(Y|X)) \]

\[ \Rightarrow \text{Var}(Y) = \text{Var}(\mathbb{E}(Y|X)) + \mathbb{E}(\text{Var}(Y|X)) \]

Exercise P. 84, 4.3

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

\[ \mathbb{P}(U \geq 1/2) = \mathbb{E}(\mathbb{1}(U \geq 1/2)|T) \]
\[ = \mathbb{E}[\mathbb{P}(U \geq 1/2|T)] \]
\[ = \mathbb{E} \left[ \frac{T - 1/2}{T} \mathbb{1}(T \geq 1/2) \right] \]
\[ = \int_{1/2}^{1} \frac{t - 1/2}{t} \, dt \]

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:

\[ S_T = X_1 + \cdots + X_T \]
\[ = X_1 1(T \geq 1) + X_2 1(T \geq 2) + \cdots \]

\[ \mathbb{E}(S_T) = \mathbb{E}(X_1 1(T \geq 1)) + \mathbb{E}(X_2 1(T \geq 2)) + \cdots \]
\[ = \mathbb{E}(X_1) + \mathbb{E}(X_2) \mathbb{P}(T \geq 2) + \mathbb{E}(X_3) \mathbb{P}(T \geq 3) + \cdots \]
\[ = \mathbb{E}(X_1) \sum_{n=1}^{\infty} \mathbb{P}(T \geq n) \]
\[ = \mathbb{E}(X_1) \mathbb{E}(T) \]

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

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

It is also necessary to justify the swap of \( \mathbb{E} \) and \( \Sigma \). 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

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

It is easy to see that \( \mathbb{E}(T) < \infty \). Just consider successive blocks if indices of length \( B + C \). 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 \( T \implies \mathbb{E}(T) < \infty \).
Let $p^+ = \mathbb{P}(S_T = +C)$ and $p^- := \mathbb{P}(S_T = -C)$. Then

\[
\mathbb{E}(S_T) = \mathbb{E}(T)\mathbb{E}(X_1) = \mathbb{E}(T) \cdot 0 = 0
\]

\[
0 = p^+ C - p^- B
\]

\[
1 = p^+ + p^-
\]

\[\implies p^+ = \frac{B}{B + C} \quad p^- = \frac{C}{B + C}\]