More on conditional expectation

Note: Math colloquium on representation of large and sparse data sets on Monday at 4 PM in Amos Eaton 214

Some advantages of working with the more abstract formalism is that some fundamental formulas become much easier to express, particularly if they are related to the coarse-graining idea.

Law of Total Expectation:

\[
E \left[ E \left[ X \mid Y \right] \right] = E \left[ X \right]
\]

(like a Fubini theorem)

Explain: average over everything, except keep \( Y \) fixed at whatever (random) value it is.

\( Y \) (only randomness left)

Conventional notation:

\[
E \left[ X \right] = \sum_{y \in Y} \left[ E \left[ X \mid Y = y \right] \right] dP_Y(y)
\]

Example:

\[
S = \sum_{n=1}^{N} X_n
\]

where \( X_n \) is independent of \( S \), \( \{X_n\}_{n=1}^{\infty} \) identically distributed, and \( N \) is independent of \( S \).
and \( N \) is independent of \( \{ X_j \}_{j=1}^\infty \):

\[
\begin{align*}
\mathbb{E}[S] &= \mathbb{E}\left[ \mathbb{E}[S | N] \right] \\
\mathbb{E}[S | N] &= \mathbb{E}\left[ \sum_{n=1}^N X_n | N \right] \\
&= \mathbb{E}\left[ \sum_{n=1}^\infty X_n \mathbb{1}\{n \leq N\} | N \right] \\
&= \sum_{n=1}^\infty \mathbb{E}[X_n \mathbb{1}\{n \leq N\} | N] \\
&= \sum_{n=1}^\infty \mathbb{1}\{n \leq N\} \mathbb{E}[X_n | N] \\
&= \sum_{n=1}^\infty \mathbb{1}\{n \leq N\} \mathbb{E}[X_n | \Xi] \\
&= \sum_{n=1}^\infty \mathbb{1}\{n \leq N\} \mathbb{E}[X_n | \Xi] \\
&= \sum_{n=1}^\infty \mathbb{1}\{n \leq N\} \mu \\
&= \mu \sum_{n=1}^N \mathbb{1}\{n \leq N\} = N \mu
\end{align*}
\]
Another fundamental law for coarse-graining

\[ \text{Var } \overline{X} = ? \] Can I also calculate this in two steps like we did for the expectation of the random sum above? Yes.

Definition of conditional variance:

\[ \text{Var } (X | Y) = \mathbb{E} \left( (X - \mathbb{E}(X | Y))^2 \bigg/ Y \right) \]

(variability of \( X \) once the r.v. \( Y \) is specified.)

Law of Total Variance:

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

Proof:

\[ \text{Var } (X) = \mathbb{E} \left( (X - \mathbb{E} X)^2 \right) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \]

(FOIL)

\[ = \mathbb{E} \left( \mathbb{E}(X^2) | Y \right) - \left( \mathbb{E} \left( \mathbb{E}(X | Y) \right) \right)^2 \]

(law of total exp.)
Lemma:
\[ \text{Var}(X | Y) = \left[ \mathbb{E}(X^2 | Y) - (\mathbb{E}(X | Y))^2 \right] \]

Proof of Lemma:
\[ \text{Var}(X | Y) = \mathbb{E}\left[ X^2 - 2XZ + Z^2 | Y \right] \]

where \( Z = \mathbb{E}[X | Y] \)

is measurable w.r.t. \( \mathcal{F}(Y) \)
\[ = \mathbb{E}[X^2 | Y] - 2Z \mathbb{E}[X | Y] + Z^2 \]
\[ = \mathbb{E}[X^2 | Y] - Z^2 \]

Lemma proved.

Now use lemma:
\[ \text{Var}(X) = \mathbb{E}\left( \text{Var}(X | Y) + Z^2 \right) \]
\[ - (\mathbb{E}(Z))^2 \]

where again \( Z = \mathbb{E}[X | Y] \)
\[ \text{Var}(X) = \mathbb{E}\left( \text{Var}(X | Y) \right) \]
\[ + \mathbb{E} Z^2 - (\mathbb{E} Z)^2 \]
\[ = \mathbb{E}\left( \text{Var}(X | Y) \right) + \text{Var}(Z) \]

This proves Law of Total Variance.
Filtrations of sigma-algebras and connections to stochastic processes

Suppose we are given a stochastic process $X(t, u)$

Define $\mathcal{F}_t$ as the $\sigma$-algebra generated by cylinder sets.
Define $\mathcal{F}_t$ as the $\sigma$-algebra generated by cylinder sets
\[ \{ w \in \Omega : X(t_i) \in B_i, \ i = 1, \ldots, n \} \]
for $t_1 \leq t_2 \leq \cdots \leq t_n \leq t$ and $\{ B_i \} \in \mathcal{B}$.

$\mathcal{F}_t^X$ is the sigma-algebra corresponding to the info generated by the stochastic process $X$ up through time $t$.

This gives a family of sigma-algebras parameterized by $t$, and they have the following relationships:

$\mathcal{F}_s^X \subseteq \mathcal{F}_t^X$ when $s \leq t$

$X(t_j \cdot)$ is measurable w.r.t. $\mathcal{F}_t^X$ for all $t$.

More generally, suppose we are given a family of $\sigma$-algebras on the prob. space $\Omega$, $s \geq t$.

\begin{align*}
& a) \quad \mathcal{F}_s \subseteq \mathcal{F}_t \quad \text{for } 0 \leq s \leq t \\
& b) \quad \mathcal{F}_t = \bigcap_{\varepsilon \to 0^+} \mathcal{F}_{t+\varepsilon} \quad \text{(right continuous)} \\
& c) \quad \mathcal{F}_t \text{ is a complete } \sigma\text{-algebra for all } t \geq 0.
\end{align*}
The notion becomes more relevant when we talk about multiple stochastic processes operating in the same probability space, and then one may ask whether one stochastic process $X(t, w)$ has the property that $X(t, \cdot)$ is measurable w.r.t. $\mathcal{F}_t$ then we say the stochastic process $X(t, w)$ is adapted to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$.

$X(t, w)$ is always adapted to
$$\{ \mathcal{F}_t \}_{t \geq 0}.$$

The notion becomes more relevant when we talk about multiple stochastic processes operating in the same probability space, and then one may ask whether one stochastic process $X(t)$ is adapted to the filtration generated by another stochastic process $Z(t)$.

Given $Z(t, w), \mathcal{F}_t$

$$X(t, w) = \int_0^t Z(s, w) \, ds \quad \text{is adapted to} \quad \mathcal{F}_t$$

$$\mathcal{Y}(t, w) = \int_0^\infty Z(s, w) \phi(t-s) \, ds \quad \text{not adapted to}$$
For the filtered process to be adapted to the filtration generated by the original process, the filter would have to be "causal"; only filter information from the past:

\[
\mathcal{Y}(t,w) = \int_0^t \mathcal{Z}(s,w) \psi(t-s) \, ds
\]

Markov property revisited:

Let \( \mathcal{F}_t^X \) be a \( \sigma \)-algebra generated by cylinder sets:

\[
\mathcal{F}_t^X = \{ \mathcal{F}(t_i) \in \mathcal{B}_i, \, i = 1, \ldots, n \}
\]

for \( t \leq t_1 \leq t_2 \leq \ldots \leq t_n \)

and \( \{ \mathcal{B}_i \} \in \mathcal{B} \)

This is the sigma algebra corresponding to information generated by the behavior of the stochastic process \( X \) at times \( t \) and later. "The future of \( X \)"

Markov property:

For any \( A \in \mathcal{F}_t^X \)
If the above statement holds true for a stochastic process $X(t)$ which is adapted to the filtration $\mathcal{F}_t$, then we say that $X(t)$ has the Markov property with respect to the filtration $\mathcal{F}_t$.

Given the present state of the stochastic process, the future of the stochastic process is independent of the past information (as defined by the filtration).