WIKIPEDIA

Law of total expectation

The proposition in probability theory known as the **law of total expectation**,[1] the **law of iterated expectations**[2] (**LIE**), the **tower rule**,[3] **Adam's law**, and the **smoothing theorem**,[4] among other names, states that if X is a random variable whose expected value $\operatorname{E}(X)$ is defined, and Y is any random variable on the same probability space, then

$$\operatorname{E}(X) = \operatorname{E}(\operatorname{E}(X \mid Y)),$$

i.e., the expected value of the conditional expected value of X given Y is the same as the expected value of X.

One special case states that if $\{A_i\}_i$ is a finite or countable partition of the sample space, then

$$\operatorname{E}(X) = \sum_i \operatorname{E}(X \mid A_i) \operatorname{P}(A_i).$$

Contents

Example

Suppose that two factories supply light bulbs to the market. Factory X's bulbs work for an average of 5000 hours, whereas factory Y's bulbs work for an average of 4000 hours. It is known that factory X supplies 60% of the total bulbs available. What is the expected length of time that a purchased bulb will work for?

Applying the law of total expectation, we have:

$$\operatorname{E}(L) = \operatorname{E}(L \mid X) \operatorname{P}(X) + \operatorname{E}(L \mid Y) \operatorname{P}(Y) = 5000(0.6) + 4000(0.4) = 4600$$

where

- $\operatorname{E}(L)$ is the expected life of the bulb;
- $\operatorname{P}(X) = \frac{6}{10}$ is the probability that the purchased bulb was manufactured by factory X;
- $\operatorname{P}(Y) = \frac{4}{10}$ is the probability that the purchased bulb was manufactured by factory Y;
- $\operatorname{E}(L \mid X) = 5000$ is the expected lifetime of a bulb manufactured by X;

- $\operatorname{E}(L \mid Y) = 4000$ is the expected lifetime of a bulb manufactured by Y.

Thus each purchased light bulb has an expected lifetime of 4600 hours.

Proof in the finite and countable cases

Let the random variables X and Y, defined on the same probability space, assume a finite or countably infinite set of finite values. Assume that $\operatorname{E}[X]$ is defined, i.e. $\min(\operatorname{E}[X_+], \operatorname{E}[X_-]) < \infty$. If $\{A_i\}$ is a partition of the probability space Ω, then

$$\operatorname{E}(X) = \sum_i \operatorname{E}(X \mid A_i) \operatorname{P}(A_i).$$

Proof.

$$\begin{aligned}
\operatorname{E}(\operatorname{E}(X \mid Y)) &= \operatorname{E}\left[\sum_x x \cdot \operatorname{P}(X = x \mid Y)\right] \\
&= \sum_y \left[\sum_x x \cdot \operatorname{P}(X = x \mid Y = y)\right] \cdot \operatorname{P}(Y = y) \\
&= \sum_y \sum_x x \cdot \operatorname{P}(X = x, Y = y).
\end{aligned}$$

If the series is finite, then we can switch the summations around, and the previous expression will become

$$\begin{aligned}
\sum_x \sum_y x \cdot \operatorname{P}(X = x, Y = y) &= \sum_x x \sum_y \operatorname{P}(X = x, Y = y) \\
&= \sum_x x \cdot \operatorname{P}(X = x) \\
&= \operatorname{E}(X).
\end{aligned}$$

If, on the other hand, the series is infinite, then its convergence cannot be conditional, due to the assumption that $\min(\operatorname{E}[X_+], \operatorname{E}[X_-]) < \infty$. The series converges absolutely if both $\operatorname{E}[X_+]$ and $\operatorname{E}[X_-]$ are finite, and diverges to an infinity when either $\operatorname{E}[X_+]$ or $\operatorname{E}[X_-]$ is infinite. In both scenarios, the above summations may be exchanged without affecting the sum.

Proof in the general case

Let $(\Omega, \mathcal{F}, \operatorname{P})$ be a probability space on which two sub σ-algebras $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$ are defined. For a random variable X on such a space, the smoothing law states that if $\operatorname{E}[X]$ is defined, i.e. $\min(\operatorname{E}[X_+], \operatorname{E}[X_-]) < \infty$, then

$$\operatorname{E}[\operatorname{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \operatorname{E}[X \mid \mathcal{G}_1] \quad \text{(a.s.)}.$$

Proof. Since a conditional expectation is a Radon–Nikodym derivative, verifying the following two properties establishes the smoothing law:

- $\operatorname{E}[\operatorname{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]$ is $\mathcal{G}_1$-measurable
- $\int_{G_1} \operatorname{E}[\operatorname{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] d\operatorname{P} = \int_{G_1} X d\operatorname{P}$, for all $G_1 \in \mathcal{G}_1$.

The first of these properties holds by definition of the conditional expectation. To prove the second one,

$$\begin{aligned}\min\left(\int_{G_1} X_+\, d\operatorname{P}, \int_{G_1} X_-\, d\operatorname{P}\right) &\leq \min\left(\int_\Omega X_+\, d\operatorname{P}, \int_\Omega X_-\, d\operatorname{P}\right) \\ &= \min(\operatorname{E}[X_+], \operatorname{E}[X_-]) < \infty,\end{aligned}$$

so the integral $\textstyle\int_{G_1} X\, d\operatorname{P}$ is defined (not equal $\infty - \infty$).

The second property thus holds since $G_1 \in \mathcal{G}_1 \subseteq \mathcal{G}_2$ implies

$$\int_{G_1} \operatorname{E}[\operatorname{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] d\operatorname{P} = \int_{G_1} \operatorname{E}[X \mid \mathcal{G}_2] d\operatorname{P} = \int_{G_1} X d\operatorname{P}.$$

Corollary. In the special case when $\mathcal{G}_1 = \{\emptyset, \Omega\}$ and $\mathcal{G}_2 = \sigma(Y)$, the smoothing law reduces to

$$\operatorname{E}[\operatorname{E}[X \mid Y]] = \operatorname{E}[X].$$

Proof of partition formula

$$\begin{aligned}\sum_i \operatorname{E}(X \mid A_i) \operatorname{P}(A_i) &= \sum_i \int\limits_\Omega X(\omega) \operatorname{P}(d\omega \mid A_i) \cdot \operatorname{P}(A_i) \\ &= \sum_i \int\limits_\Omega X(\omega) \operatorname{P}(d\omega \cap A_i) \\ &= \sum_i \int\limits_\Omega X(\omega) I_{A_i}(\omega) \operatorname{P}(d\omega) \\ &= \sum_i \operatorname{E}(X I_{A_i}),\end{aligned}$$

where I_{A_i} is the indicator function of the set A_i.

If the partition $\{A_i\}_{i=0}^n$ is finite, then, by linearity, the previous expression becomes

$$\operatorname{E}\left(\sum_{i=0}^n X I_{A_i}\right) = \operatorname{E}(X),$$

and we are done.

If, however, the partition $\{A_i\}_{i=0}^{\infty}$ is infinite, then we use the dominated convergence theorem to show that

$$\mathrm{E}\left(\sum_{i=0}^{n} XI_{A_i}\right) \to \mathrm{E}(X).$$

Indeed, for every $n \geq 0$,

$$\left|\sum_{i=0}^{n} XI_{A_i}\right| \leq |X| I_{\bigcup\limits_{i=0}^{n} A_i} \leq |X|.$$

Since every element of the set Ω falls into a specific partition A_i, it is straightforward to verify that the sequence $\left\{\sum_{i=0}^{n} XI_{A_i}\right\}_{n=0}^{\infty}$ converges pointwise to X. By initial assumption, $\mathrm{E}\,|X| < \infty$. Applying the dominated convergence theorem yields the desired.

See also

- The fundamental theorem of poker for one practical application.
- Law of total probability
- Law of total variance
- Law of total covariance
- Law of total cumulance
- Product distribution#expectation (application of the Law for proving that the product expectation is the product of expectations)

References

1. Weiss, Neil A. (2005). *A Course in Probability* (https://books.google.com/books?id=p-rwJAAACAAJ&pg=PA380). Boston: Addison–Wesley. pp. 380–383. ISBN 0-321-18954-X.
2. "Law of Iterated Expectation | Brilliant Math & Science Wiki" (https://brilliant.org/wiki/law-of-iterated-expectation/). *brilliant.org*. Retrieved 2018-03-28.
3. Rhee, Chang-han (Sep 20, 2011). "Probability and Statistics" (https://web.stanford.edu/class/cme001/handouts/changhan/Refresher2.pdf) (PDF).
4. Wolpert, Robert (November 18, 2010). "Conditional Expectation" (https://www2.stat.duke.edu/courses/Fall10/sta205/lec/topics/rn.pdf) (PDF).

- Billingsley, Patrick (1995). *Probability and measure*. New York: John Wiley & Sons. ISBN 0-471-00710-2. (Theorem 34.4)
- Christopher Sims, "Notes on Random Variables, Expectations, Probability Densities, and Martingales" (http://sims.princeton.edu/yftp/Bubbles2007/ProbNotes.pdf), especially equations (16) through (18)