Law of the unconscious statistician

From Wikipedia, the free encyclopedia

In probability theory and statistics, the **law of the unconscious statistician**, or **LOTUS**, is a theorem which expresses the expected value of a function $g(X)$ of a random variable X in terms of g and the probability distribution of X.

The form of the law depends on the type of random variable X in question. If the distribution of X is discrete and one knows its probability mass function p_X, then the expected value of $g(X)$ is

$$\operatorname{E}[g(X)] = \sum_x g(x) p_X(x),$$

where the sum is over all possible values x of X. If instead the distribution of X is continuous with probability density function f_X, then the expected value of $g(X)$ is

$$\operatorname{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, \mathrm{d}x$$

Both of these special cases can be expressed in terms of the cumulative probability distribution function F_X of X, with the expected value of $g(X)$ now given by the Lebesgue–Stieltjes integral

$$\operatorname{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \, \mathrm{d}F_X(x).$$

In even greater generality, X could be a random element in any measurable space, in which case the law is given in terms of measure theory and the Lebesgue integral. In this setting, there is no need to restrict the context to probability measures, and the law becomes a general theorem of mathematical analysis on Lebesgue integration relative to a pushforward measure.

Etymology [edit]

This proposition is (sometimes) known as the *law of the unconscious statistician* because of a purported tendency to think of the aforementioned law as the very definition of the expected value of a function $g(X)$ and a random variable X, rather than (more formally) as a consequence of the true definition of expected value.[1] The naming is sometimes attributed to Sheldon Ross' textbook *Introduction to Probability Models*, although he removed the reference in later editions.[2] Many statistics textbooks do present the result as the definition of expected value.[3]

Joint distributions [edit]

A similar property holds for joint distributions, or equivalently, for random vectors. For discrete random variables X and Y, a function of two variables g, and joint probability mass function $p_{X,Y}(x,y)$:[4]

$$\operatorname{E}[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y)$$

In the absolutely continuous case, with $f_{X,Y}(x,y)$ being the joint probability density function,

$$\operatorname{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y$$

Special cases [edit]

A number of special cases are given here. In the simplest case, where the random variable X takes on countably many values (so that its distribution is discrete), the proof is particularly simple, and holds without modification if X is a discrete random vector or even a discrete random element.

The case of a continuous random variable is more subtle, since the proof in generality requires subtle forms of the change-of-variables formula for integration. However, in the framework of measure theory, the discrete case generalizes straightforwardly to general (not necessarily discrete) random elements, and the case of a continuous random variable is then a special case by making use of the Radon–Nikodym theorem.

Discrete case [edit]

Suppose that X is a random variable which takes on only finitely or countably many different values $x_1, x_2, \ldots$, with probabilities $p_1, p_2, \ldots$. Then for any function g of these values, the random variable $g(X)$ has values $g(x_1), g(x_2), \ldots$, although some of these may coincide with each other. For example, this is the case if X can take on both values 1 and -1 and $g(x) = x^2$.

Let $y_1, y_2, \ldots$ enumerate the possible *distinct* values of $g(X)$, and for each i let I_i denote the collection of all j with $g(x_j) = y_i$. Then, according to the definition of expected value, there is

$$\operatorname{E}[g(X)] = \sum_i y_i p_{g(X)}(y_i).$$

Since a y_i can be the image of multiple, distinct x_j, it holds that

$$p_{g(X)}(y_i) = \sum_{j \in I_i} p_X(x_j).$$

Then the expected value can be rewritten as

$$\sum_i y_i p_{g(X)}(y_i) = \sum_i y_i \sum_{j \in I_i} p_X(x_j) = \sum_i \sum_{j \in I_i} g(x_j) p_X(x_j) = \sum_x g(x) p_X(x).$$ This equality relates the average of the outputs of $g(X)$ as weighted by the probabilities of the outputs themselves to the average of the outputs of $g(X)$ as weighted by the probabilities of the outputs of X.

If X takes on only finitely many possible values, the above is fully rigorous. However, if X takes on countably many values, the last equality given does not always hold, as seen by the Riemann series theorem. Because of this, it is necessary to assume the absolute convergence of the sums in question.[5]

Continuous case [edit]

Suppose that X is a random variable whose distribution has a continuous density f. If g is a general function, then the probability that $g(X)$ is valued in a set of real numbers K equals the probability that X is valued in $g^{-1}(K)$, which is given by $$\int_{g^{-1}(K)} f(x)\,\mathrm{d}x.$$ Under various conditions on g, the change-of-variables formula for integration can be applied to relate this to an integral over K, and hence to identify the density of $g(X)$ in terms of the density of X. In the simplest case, if g is differentiable with nowhere-vanishing derivative, then the above integral can be written as $$\int_K f(g^{-1}(y))(g^{-1})'(y)\,\mathrm{d}y,$$ thereby identifying $g(X)$ as possessing the density $f(g^{-1}(y))(g^{-1})'(y)$. The expected value of $g(X)$ is then identified as

$$\int_{-\infty}^{\infty} y f(g^{-1}(y))(g^{-1})'(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x,$$ where the equality follows by another use of the change-of-variables formula for integration. This shows that the expected value of $g(X)$ is encoded entirely by the function g and the density f of X.[6]

The assumption that g is differentiable with nonvanishing derivative, which is necessary for applying the usual change-of-variables formula, excludes many typical cases, such as $g(x) = x^2$. The result still holds true in these broader settings, although the proof requires more sophisticated results from mathematical analysis such as Sard's theorem and the coarea formula. In even greater generality, using the Lebesgue theory as below, it can be found that the identity $$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x$$ holds true whenever X has a density f (which does not have to be continuous) and whenever g is a measurable function for which $g(X)$ has finite expected value. (Every continuous function is measurable.) Furthermore, without modification to the proof, this holds even if X is a random vector (with density) and g is a multivariable function; the integral is then taken over the multi-dimensional range of values of X.

Measure-theoretic formulation [edit]

An abstract and general form of the result is available using the framework of measure theory and the Lebesgue integral. Here, the setting is that of a measure space (Ω, μ) and a measurable map X from Ω to a measurable space Ω'. The theorem then says that for any measurable function g on Ω' which is valued in real numbers (or even the extended real number line), there is

$$\int_{\Omega} g \circ X \, \mathrm{d}\mu = \int_{\Omega'} g \, \mathrm{d}(X_\sharp \mu),$$

(interpreted as saying, in particular, that either side of the equality exists if the other side exists). Here $X_\sharp \mu$ denotes the pushforward measure on Ω'. The 'discrete case' given above is the special case arising when X takes on only countably many values and μ is a probability measure. In fact, the discrete case (although without the restriction to probability measures) is the first step in proving the general measure-theoretic formulation, as the general version follows therefrom by an application of the monotone convergence theorem.[7] Without any major changes, the result can also be formulated in the setting of outer measures.[8]

If μ is a σ-finite measure, the theory of the Radon–Nikodym derivative is applicable. In the special case that the measure $X_\sharp \mu$ is absolutely continuous relative to some background σ-finite measure ν on Ω', there is a real-valued function f_X on Ω' representing the Radon–Nikodym derivative of the two measures, and then

$$\int_{\Omega'} g \, \mathrm{d}(X_\sharp \mu) = \int_{\Omega'} g f_X \, \mathrm{d}\nu.$$

In the further special case that Ω' is the real number line, as in the contexts discussed above, it is natural to take ν to be the Lebesgue measure, and this then recovers the 'continuous case' given above whenever μ is a probability measure. (In this special case, the condition of σ-finiteness is vacuous, since Lebesgue measure and every probability measure are trivially σ-finite.)[9]

References [edit]

1. ^ DeGroot & Schervish 2014, pp. 213−214.
2. ^ Casella & Berger 2001, Section 2.2; Ross 2019.
3. ^ Casella & Berger 2001, Section 2.2.
4. ^ Ross 2019.
5. ^ Feller 1968, Section IX.2.
6. ^ Papoulis & Pillai 2002, Chapter 5.
7. ^ Bogachev 2007, Section 3.6; Cohn 2013, Section 2.6; Halmos 1950, Section 39.
8. ^ Federer 1969, Section 2.4.
9. ^ Halmos 1950, Section 39.

- Bogachev, V. I. (2007). *Measure theory. Volume I*. Berlin: Springer-Verlag. doi:10.1007/978-3-540-34514-5 . ISBN 978-3-540-34513-8. MR 2267655 . Zbl 1120.28001 .
- Casella, George; Berger, Roger L. (2001). *Statistical inference*. Duxbury Advanced Series (Second edition of 1990 original ed.). Pacific Grove, CA: Duxbury. ISBN 0-534-11958-1. Zbl 0699.62001 .
- Cohn, Donald L. (2013). *Measure theory*. Birkhäuser Advanced Texts: Basler Lehrbücher (Second edition of 1980 original ed.). New York: Birkhäuser/Springer. doi:10.1007/978-1-4614-6956-8 . ISBN 978-1-4614-6955-1. MR 3098996 . Zbl 1292.28002 .
- DeGroot, Morris H.; Schervish, Mark J. (2014). *Probability and statistics* (Fourth edition of 1975 original ed.). Pearson Education. ISBN 0-321-50046-6. MR 0373075 . Zbl 0619.62001 .

- Federer, Herbert (1969). *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften. Vol. 153. Berlin–Heidelberg–New York: Springer-Verlag. doi:10.1007/978-3-642-62010-2. ISBN 978-3-540-60656-7. MR 0257325. Zbl 0176.00801.
- Feller, William (1968). *An introduction to probability theory and its applications. Volume I* (Third edition of 1950 original ed.). New York–London–Sydney: John Wiley & Sons, Inc. MR 0228020. Zbl 0155.23101.
- Halmos, Paul R. (1950). *Measure theory*. New York: D. Van Nostrand Co., Inc. doi:10.1007/978-1-4684-9440-2. MR 0033869. Zbl 0040.16802.
- Papoulis, Athanasios; Pillai, S. Unnikrishna (2002). *Probability, random variables, and stochastic processes* (Fourth edition of 1965 original ed.). New York: McGraw-Hill. ISBN 0-07-366011-6.
- Ross, Sheldon M. (2019). *Introduction to probability models* (Twelfth edition of 1972 original ed.). London: Academic Press. doi:10.1016/C2017-0-01324-1. ISBN 978-0-12-814346-9. MR 3931305. Zbl 1408.60002.