WIKIPEDIA

Law of total variance

In probability theory, the **law of total variance**[1] or **variance decomposition formula** or **conditional variance formulas** or **law of iterated variances** also known as **Eve's law**[2], states that if X and Y are random variables on the same probability space, and the variance of Y is finite, then

$$\operatorname{Var}(Y) = \operatorname{E}[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(\operatorname{E}[Y \mid X]).$$

In language perhaps better known to statisticians than to probabilists, the two terms are the "unexplained" and the "explained" components of the variance respectively (cf. fraction of variance unexplained, explained variation). In actuarial science, specifically credibility theory, the first component is called the expected value of the process variance (**EVPV**) and the second is called the variance of the hypothetical means (**VHM**).[3]

There is a general variance decomposition formula for $c \geq 2$ components (see below).[4] For example, with two conditioning random variables:

$$\operatorname{Var}[Y] = \operatorname{E}[\operatorname{Var}(Y \mid X_1, X_2)] + \operatorname{E}[\operatorname{Var}(\operatorname{E}[Y \mid X_1, X_2] \mid X_1)] + \operatorname{Var}(\operatorname{E}[Y \mid X_1]),$$

which follows from the law of total conditional variance:[4]

$$\operatorname{Var}(Y \mid X_1) = \operatorname{E}[\operatorname{Var}(Y \mid X_1, X_2) \mid X_1] + \operatorname{Var}(\operatorname{E}[Y \mid X_1, X_2] \mid X_1).$$

Note that the conditional expected value $\operatorname{E}(Y \mid X)$ is a random variable in its own right, whose value depends on the value of X. Notice that the conditional expected value of Y given the *event* $X = x$ is a function of x (this is where adherence to the conventional and rigidly case-sensitive notation of probability theory becomes important!). If we write $\operatorname{E}(Y \mid X = x) = g(x)$ then the random variable $\operatorname{E}(Y \mid X)$ is just $g(X)$. Similar comments apply to the conditional variance.

One special case, (similar to the law of total expectation) states that if $A_1, \ldots, A_n$ is a partition of the whole outcome space, i.e. these events are mutually exclusive and exhaustive, then

$$\begin{aligned}\operatorname{Var}(X) = {} & \sum_{i=1}^{n} \operatorname{Var}(X \mid A_i) \Pr(A_i) + \sum_{i=1}^{n} \operatorname{E}[X \mid A_i]^2 (1 - \Pr(A_i)) \Pr(A_i) \\ & - 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \operatorname{E}[X \mid A_i] \Pr(A_i) \operatorname{E}[X \mid A_j] \Pr(A_j).\end{aligned}$$

In this formula, the first component is the expectation of the conditional variance; the other two rows are the variance of the conditional expectation.

Contents

Proof

General variance decomposition applicable to dynamic systems

The square of the correlation and explained (or informational) variation

Higher moments

See also

References

Proof

The law of total variance can be proved using the law of total expectation.[5] First,

$$\operatorname{Var}[Y] = \operatorname{E}[Y^2] - \operatorname{E}[Y]^2$$

from the definition of variance. Then we apply the law of total expectation to each term by conditioning on the random variable *X*:

$$= \operatorname{E}\left[\operatorname{E}[Y^2 \mid X]\right] - [\operatorname{E}[\operatorname{E}[Y \mid X]]]^2$$

Now we rewrite the conditional second moment of Y in terms of its variance and first moment:

$$= \operatorname{E}\left[\operatorname{Var}[Y \mid X] + [\operatorname{E}[Y \mid X]]^2\right] - [\operatorname{E}[\operatorname{E}[Y \mid X]]]^2$$

Since the expectation of a sum is the sum of expectations, the terms can now be regrouped:

$$= \operatorname{E}[\operatorname{Var}[Y \mid X]] + \left(\operatorname{E}[[\operatorname{E}[Y \mid X]]^2] - [\operatorname{E}[\operatorname{E}[Y \mid X]]]^2\right)$$

Finally, we recognize the terms in parentheses as the variance of the conditional expectation E[*Y* | *X*]:

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

General variance decomposition applicable to dynamic systems

The following formula shows how to apply the general, measure theoretic variance decomposition formula [4] to stochastic dynamic systems. Let *Y*(*t*) be the value of a system variable at time *t*. Suppose we have the internal histories (natural filtrations) $H_{1t}, H_{2t}, \ldots, H_{c-1,t}$, each one corresponding to the history (trajectory) of a different collection of system variables. The collections need not be disjoint. The variance of *Y*(*t*) can be decomposed, for all times *t*, into *c* ≥ 2 components as follows:

$$\begin{aligned}\operatorname{Var}[Y(t)] = {} & \operatorname{E}(\operatorname{Var}[Y(t) \mid H_{1t}, H_{2t}, \ldots, H_{c-1,t}]) \\ & + \sum_{j=2}^{c-1} \operatorname{E}(\operatorname{Var}[\operatorname{E}[Y(t) \mid H_{1t}, H_{2t}, \ldots, H_{jt}] \mid H_{1t}, H_{2t}, \ldots, H_{j-1,t}]) \\ & + \operatorname{Var}(\operatorname{E}[Y(t) \mid H_{1t}]).\end{aligned}$$

The decomposition is not unique. It depends on the order of the conditioning in the sequential decomposition.

The square of the correlation and explained (or informational) variation

In cases where (*Y*, *X*) are such that the conditional expected value is linear; i.e., in cases where

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

it follows from the bilinearity of covariance that

$$a = \frac{\operatorname{Cov}(Y,X)}{\operatorname{Var}(X)}$$

and

$$b = \operatorname{E}(Y) - \frac{\operatorname{Cov}(Y,X)}{\operatorname{Var}(X)} \operatorname{E}(X)$$

and the explained component of the variance divided by the total variance is just the square of the correlation between *Y* and *X*; i.e., in such cases,

$$\frac{\operatorname{Var}(\operatorname{E}(Y \mid X))}{\operatorname{Var}(Y)} = \operatorname{Corr}(X,Y)^2.$$

One example of this situation is when (*X*, *Y*) have a bivariate normal (Gaussian) distribution.

More generally, when the conditional expectation E(*Y* | *X*) is a non-linear function of *X*

$$\iota_{Y|X} = \frac{\operatorname{Var}(\operatorname{E}(Y \mid X))}{\operatorname{Var}(Y)} = \operatorname{Corr}(\operatorname{E}(Y \mid X), Y)^2,$$ [4]

which can be estimated as the *R* squared from a non-linear regression of *Y* on *X*, using data drawn from the joint distribution of (*X*,*Y*). When E(*Y* | *X*) has a Gaussian distribution (and is an invertible function of *X*), or *Y* itself has a (marginal) Gaussian distribution, this explained component of variation sets a lower bound on the mutual information:[4]

$$\operatorname{I}(Y;X) \geq \ln([1 - \iota_{Y|X}]^{-1/2}).$$

Higher moments

A similar law for the third central moment μ_3 says

$$\mu_3(Y) = \operatorname{E}(\mu_3(Y \mid X)) + \mu_3(\operatorname{E}(Y \mid X)) + 3\operatorname{cov}(\operatorname{E}(Y \mid X), \operatorname{var}(Y \mid X)).$$

For higher cumulants, a generalization exists. See law of total cumulance.

See also

- Law of total covariance, a generalization
- Law of propagation of errors

References

1. Neil A. Weiss, *A Course in Probability*, Addison–Wesley, 2005, pages 385–386.
2. Joseph K. Blitzstein and Jessica Hwang: "Introduction to Probability"
3. Mahler, Howard C.; Dean, Curtis Gary (2001). "Chapter 8: Credibility" (http://people.stat.sfu.ca/~cltsai/ACMA315/Ch8_Credibility.pdf) (PDF). In Casualty Actuarial Society (ed.). *Foundations of Casualty Actuarial Science* (4th ed.). Casualty Actuarial Society. pp. 525–526. ISBN 978-0-96247-622-8. Retrieved June 25, 2015.
4. Bowsher, C.G. and P.S. Swain, Proc Natl Acad Sci USA, 2012: 109, E1320–29.
5. Neil A. Weiss, *A Course in Probability*, Addison–Wesley, 2005, pages 380–383.

- Blitzstein, Joe. "Stat 110 Final Review (Eve's Law)" (http://projects.iq.harvard.edu/files/stat110/files/final_review.pdf) (PDF). *stat110.net*. Harvard University, Department of Statistics. Retrieved 9 July 2014.
- Billingsley, Patrick (1995). *Probability and Measure*. New York, NY: John Wiley & Sons, Inc. ISBN 0-471-00710-2. (Problem 34.10(b))