
Dos expresiones para la esperanza de una variable aleatoria

Two Expressions about the Expectation of a Random Variable

Andrés Gutiérrez^a
hugogutierrez@usantotomas.edu.co

Hanwen Zhang^b
predictive@telmex.net.co

Resumen

La definición de esperanza de una variable aleatoria en los textos clásicos es diferente a la definición que se da en un curso de probabilidad basado en la teoría de la medida. En este manuscrito se muestra por qué, efectivamente, estas dos definiciones coinciden.

Palabras clave: esperanza, variable aleatoria.

Abstract

In the basic texts, the definition of the expected value of a random variable is different from the definition given in a probability course based on the theory of measure. In this manuscript it is shown why both definitions are in fact equivalent.

Key words: expected value, random variable.

1. Introducción

La esperanza de una variable aleatoria, uno de los conceptos más importantes y poderosos de la teoría estadística, parecería tener dos definiciones distintas, dependiendo del nivel académico en que sea vista. Por una parte, está la definición desde el punto de vista de la teoría de la medida. Por ejemplo, Shao (2003, p. 11) afirma que la *esperanza* o *valor esperado* de una variable aleatoria X está dado por la siguiente expresión:

$$E(X) = \int_{\Omega} X dP \quad (1.1)$$

^aDirector. Centro de Investigaciones y Estudios Estadísticos (CIEES)

^bInvestigador. Predictive Ltda.

donde P es la medida de probabilidad del espacio de referencia¹. Sin embargo, la anterior definición difiere con la de un libro de texto clásico de probabilidad como Mood et al. (1963, p. 69), en donde la esperanza de una variable aleatoria continua está dada por la siguiente expresión:

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx \quad (1.2)$$

Es claro que las dos definiciones no concuerdan a simple vista. En este manuscrito se aclara por qué, en efecto, las dos definiciones coinciden.

1.1. Resultados útiles

Para llevar a cabo la demostración de la coincidencia de las dos definiciones de esperanza, es útil hacer un breve repaso por los resultados de la teoría de la media y más específicamente una medida de probabilidad.

Definición 1.1. *Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad. Una variable aleatoria X es una función medible de (Ω, \mathfrak{F}) en $(\mathbb{R}, \mathfrak{B})$. Donde \mathfrak{B} es la σ -álgebra de Borel.*

Definición 1.2. *Una variable aleatoria X induce una medida de probabilidad μ_X sobre \mathbb{R} tal que*

$$\mu_X = P \circ X^{-1} = P(X^{-1}(A)) = P(\omega \in \Omega : X(\omega) \in A) \quad (1.3)$$

Nótese que el espacio de llegada de una variable aleatoria puede ser medible para diferentes medidas. Sin embargo, existe sólo una medida de probabilidad que cumple la anterior definición y que está ligada de forma estructural a la variable aleatoria. Como ilustración, el siguiente diagrama muestra cómo la variable aleatoria X induce a la medida de probabilidad μ_X .

$$X : (\Omega, \mathfrak{F}, P) \longrightarrow (\mathbb{R}, \mathfrak{B}, \mu_X := P \circ X^{-1})$$

Definición 1.3. *Sea $A \subset \Omega$ un conjunto medible. La función indicadora para A se define como*

$$I_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A; \\ 0, & \text{en otro caso.} \end{cases} \quad (1.4)$$

Definición 1.4. *Una función simple φ es una combinación lineal de las funciones indicadoras de conjuntos medibles, definida como*

$$\varphi(\omega) = \sum_{i=1}^k a_i I_{A_i}(\omega) \quad (1.5)$$

¹Nótese que esta definición es general para cualquier variable aleatoria, sea continua o discreta. El carácter de la integral (o sumatoria) lo determina la medida del espacio de referencia.

Resultado 1.1. Sea f una función de Borel no negativa sobre (Ω, \mathfrak{F}) . Entonces existe una sucesión de funciones simples $\{\varphi_n\}$ las cuales satisfacen que $0 \leq \varphi_1 \leq \varphi_2 \leq \dots \leq \varphi_n \leq g$ y

$$\lim_{n \rightarrow \infty} \varphi_n = g \tag{1.6}$$

Definición 1.5. Para cualquier función de Borel f , se define la parte positiva como

$$f_+(\omega) = \text{máx}\{f(\omega), 0\} \tag{1.7}$$

y la parte negativa como

$$f_-(\omega) = \text{máx}\{-f(\omega), 0\} \tag{1.8}$$

Nótese que $f(\omega) = f_+(\omega) + f_-(\omega)$.

Definición 1.6. Sea $A \subset \Omega$ un conjunto medible y f una función de Borel no negativa. La integral de f sobre A se define como

$$\int_A f d\mu = \int_{\Omega} I_A f d\mu \tag{1.9}$$

Resultado 1.2. (Linealidad de las integrales) Sea $(\Omega, \mathfrak{F}, \mu)$ un espacio de medida y f y g funciones de Borel integrable.

- $\int (af) d\mu = a \int f d\mu$, con $a \in \mathbb{R}$.
- $\int (f + g) d\mu = \int f d\mu + \int g d\mu$.

Resultado 1.3. (Convergencia monótona) Sea f_1, f_2, \dots una sucesión de funciones de Borel sobre $(\Omega, \mathfrak{F}, \mu)$. Si $0 \leq f_1 \leq f_2 \leq \dots$ y además $\lim_{n \rightarrow \infty} f_n = f$, entonces

$$\int \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu \tag{1.10}$$

Definición 1.7. La integral de una función simple no negativa φ con respecto a la medida μ está dada por

$$\int \varphi d\mu = \sum_{i=1}^k a_i \mu(A_i) \tag{1.11}$$

como caso particular, cuando la función simple consta de un sólo término tal que $a_1 = 1$, entonces la integral de esta función con respecto a la medida μ está dada por

$$\int I(A) d\mu = \mu(A) \tag{1.12}$$

Resultado 1.4. (Teorema de Radon-Nikodym) Sea m la medida de Lebesgue. Si μ_X es absolutamente continua con respecto a m , entonces existe una función de Borel no negativa $f_X : \mathbb{R} \rightarrow \mathbb{R}$ tal que

$$\mu_X(A) = \int_A f_X dm \tag{1.13}$$

La función f es una función de densidad de probabilidad para la variable aleatoria X . Nótese que la integral de Lebesgue de f_X sobre un intervalo (a, b) coincide con la integral de Riemann definida como

$$\int_a^b f_X(x) dx \quad (1.14)$$

2. Coincidencia

Cook (2008) afirma que el siguiente teorema es llamado “**la ley del estadístico inconsciente**” puesto que este es aplicado tan frecuentemente que se hace de manera inconsciente e indiferente.

Resultado 2.1. Si $g : \mathbb{R} \rightarrow \mathbb{R}$ es una función μ_X -medible, entonces $g(X)$ es P -medible y además

$$\int_{\Omega} g(X) dP = \int_{\mathbb{R}} g(x) f_X(x) dx \quad (2.1)$$

Si X tiene una función de densidad de probabilidad $f_X(\cdot)$, entonces

$$E(g(X)) = \int_{\mathbb{R}} g(x) f_X(x) dx \quad (2.2)$$

Como caso particular, cuando la función $g(x) = x$, entonces la esperanza de la variable aleatoria X toma la forma clásica dada por (1.1).

Demostración. La demostración de este teorema se realiza en cuatro pasos, para cuatro funciones (figura 1). En primer lugar se verifica el teorema para una función indicadora, luego se prueba que el teorema aplica para cualquier función simple y por consiguiente también para cualquier función no negativa. En último lugar, se demuestra que, como consecuencia de lo anterior, el teorema se verifica para cualquier función $g(\cdot)$.

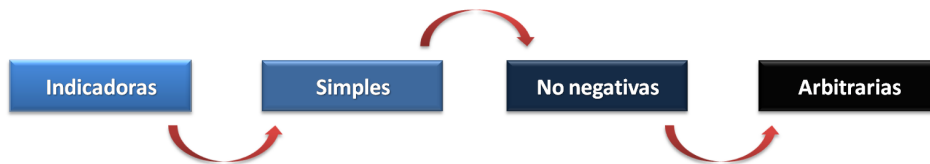


Figura 1: Pasos para la demostración del teorema.

1. **Función indicadora:** sea A un conjunto de Borel, tal que $A \subseteq \mathbb{R}$. Luego

$$\begin{aligned} \int_{\mathbb{R}} I_A(x) f_X(x) dx &= \int_A f_X(x) dx \\ &= \mu_X(A) \\ &= P(X^{-1}(A)) \end{aligned}$$

En donde se utiliza la definición de integral dada por (1.9), el resultado de Radon-Nikodyn y la definición de medida inducida por una variable.

Por otro lado, definiendo $S = \{\omega \in \Omega : X(\omega) \in A\}$, se tiene que

$$\begin{aligned} \int_{\Omega} I_A(X(\omega)) dP &= \int_S I_A(X(\omega)) dP + \int_{S^c} I_A(X(\omega)) dP \\ &= \int_S 1 dP + \int_{S^c} 0 dP \\ &= \int_S I_S(\omega) dP \\ &= P(\omega \in \Omega : X(\omega) \in A) = P(X^{-1}(A)) \end{aligned}$$

En donde se utiliza el hecho de que la integral de una función indicadora es la medida del conjunto medible. Nótese que la cadena de igualdades coincide con la anterior y por consiguiente se tiene el cumplimiento del teorema puesto que

$$\int_{\mathbb{R}} I_A(x) f_X(x) dx = \int_{\Omega} I_A(X(\omega)) dP \quad (2.3)$$

2. **Función simple:** como el teorema es válido para funciones indicadoras y utilizando la linealidad de las integrales, se tiene que

$$\begin{aligned} \int_{\mathbb{R}} \varphi(x) f_X(x) dx &= \int_{\mathbb{R}} \sum_{i=1}^k a_i I_{A_i}(x) f_X(x) dx \\ &= \sum_{i=1}^k a_i \int_{\mathbb{R}} I_{A_i}(x) f_X(x) dx \\ &= \sum_{i=1}^k a_i \int_{\Omega} I_{A_i}(X(\omega)) dP \\ &= \int_{\Omega} \sum_{i=1}^k a_i I_{A_i}(X(\omega)) dP \\ &= \int_{\Omega} \varphi(X(\omega)) dP \end{aligned}$$

3. **Función no negativa:** dado que f_X es la derivada de Radon-Nikodyn, entonces es no negativa. Ahora, utilizando (2.4), se tiene que existe una

sucesión de funciones $\{f_x \varphi_n\}$ las cuales satisfacen que $0 \leq f_x \varphi_1 \leq f_x \varphi_2 \leq \dots \leq f_x \varphi_n \leq f_x g$ y

$$\lim_{n \rightarrow \infty} f_x \varphi_n = f_x g \quad (2.4)$$

De esta forma, aplicando el resultado de convergencia monótona, se tiene que

$$\begin{aligned} \int_{\Omega} g(X(\omega)) dP &= \int_{\Omega} \lim_{n \rightarrow \infty} \varphi_n(X(\omega)) dP \\ &= \lim_{n \rightarrow \infty} \int_{\Omega} \varphi_n(X(\omega)) dP \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \varphi(x) f_X(x) dx \\ &= \int_{\mathbb{R}} \lim_{n \rightarrow \infty} \varphi(x) f_X(x) dx \\ &= \int_{\mathbb{R}} g(x) f_X(x) dx \end{aligned}$$

con lo cual se tiene que el teorema es válido para cualquier función no negativa g .

4. **Función arbitraria:** aplicando la definición de parte negativa y positiva de una función de Borel, se tiene que $g(w) = g_+(w) - g_-(w)$, con g_+ y g_- funciones no negativas. Entonces aplicando el anterior resultado, se tiene

$$\begin{aligned} \int_{\Omega} g_+(X(\omega)) dP &= \int_{\mathbb{R}} g_+(x) f_X(x) dx \\ \int_{\Omega} g_-(X(\omega)) dP &= \int_{\mathbb{R}} g_-(x) f_X(x) dx \end{aligned} \quad (2.5)$$

Por consiguiente

$$\begin{aligned} \int_{\Omega} g_+(X(\omega)) dP - \int_{\Omega} g_-(X(\omega)) dP &= \\ &= \int_{\mathbb{R}} g_+(x) f_X(x) dx - \int_{\mathbb{R}} g_-(x) f_X(x) dx \end{aligned}$$

Por la linealidad de las integrales, se tiene que

$$\int_{\Omega} [g_+ - g_-](X(\omega)) dP = \int_{\mathbb{R}} [g_+ - g_-](x) f_X(x) dx$$

Y con esto se tiene la demostración completa del teorema puesto que

$$\int_{\Omega} g(X(\omega)) dP = \int_{\mathbb{R}} g(x) f_X(x) dx \quad \square$$

Nótese que cuando la función g es la función idéntica, se tiene que las dos definiciones de esperanza coinciden. Es decir

$$E(X) = \int_{\mathbb{R}} xf_X(x) dx = \int_{\Omega} X dP \quad (2.6)$$

3. Discusión

En este manuscrito se ha probado que, efectivamente, las dos definiciones de esperanza coinciden. Existen otras formas de probar este resultado; sin embargo, aquí quisimos hacerla de una forma clara, de acuerdo a los procedimientos clásicos de la teoría de la medida.

Nótese que la igualdad de las dos esperanzas es una consecuencia de la ley del estadístico inconsciente dada por el resultado 2.1. Cuando la función g se toma distinta a la función identidad, se pueden probar otro tipo de igualdades que resultan interesantes en el desarrollo de la teoría estadística; por ejemplo, que las dos definiciones de la función generadora de momento son iguales, o la definición de varianza.

Agradecimientos

A los lectores de *Apuntes de Estadística* (Apuntes de Estadística 2008) por sus invaluable comentarios. A Camilo Cubides por su colaboración en la corrección del estilo de este artículo.

Referencias

Apuntes de Estadística (2008), *El teorema del estadístico inconsciente*.

*<http://www.predictive.wordpress.com>

Cook, J. (2008), 'Relating two Definitions of Expectations', *The Endeavour blog* .

*<http://www.johndcook.com/blog/>

Mood, A. M., Graybill, F. & Boes, D. C. (1963), *Introduction to the Theory of Statistics*, McGraw-Hill.

Shao, J. (2003), *Mathematical Statistics*, Springer.