Theory of Conditional Expectation

Piyush Ahuja

***Abstract*—This paper discusses the theory of Conditional Expectation, which will help us in evaluating a random variable to the best of our knowledge i.e with the information we have. The properties required of such a function will motivate its definition, which is also found i n many undergraduate textbooks. The existence of Conditional Expectation will then be proved from Radon-Nikodym Theorem. In the later half, there will be an attempt to provide an attempt to provide an alternative view of Conditional Expectation. It will be shown that the existence of such a function arises naturally out of the structure of the Hilbert Space $L_2(\Omega, \mathcal{F}, \mathbb{P})$, and this will be proved without using the Radon-Nikodym Theorem. The paper will be rounded off by proving some properties arising naturally as a result of our definition, which will be found useful in Financial Applications.**

***Index Terms*—Radon-Nikodym, Conditional Expectation, Measure Space, Hilbert Space, Banach Space, Orthogonal Projection**

I. Preliminaries

A measure is the generalisation of length.

Definition 1. A measure is an extended real-valued function μ defined on a $\sigma - algebra$ $\mathbb{X}$ of subsets of X such that :

i) $\mu(\emptyset)$ = 0 ii) $\mu(E) \geqslant 0 \; \forall \; E \in \mathbb{X}$ iii) $\mu$is countably additive in the sense that if (E_n) is any disjoint sequence of sets in $\mathbb{X}$, then $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$

A measure space is a triple $(X, \mathbb{X}, \mu)$ consisting of a set X, a $\sigma - algebra$ $\mathbb{X}$of subsets of X, and a measure μdefined on $\mathbb{X}$.

II. Motivation

Consider a random variable defined on S defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a $sub - \sigma - algebra$ $\mathcal{G}$ of $\mathcal{F}$. If S is $\mathcal{G}$-measurable, then the information in $\mathcal{G}$ is sufficient to determine the value of S. If S is independent of $\mathcal{G}$, then the information in $\mathcal{G}$is sufficient to determine the value of S. If S is independent of $\mathcal{G}$, then the information in $\mathcal{G}$ provides no help in evaluating S. In the intermediate case, we can use the information in $\mathcal{G}$ to estimate but not precisely evaluate S. We'll call the estimate of S based on the information in $\mathcal{G}$ the conditional expectation of S given $\mathcal{G}$, denoted by $\mathbb{E}[S|\mathcal{G}]$.

Note that we need the following properties from $\mathbb{E}[S|\mathcal{G}]$:

- $\mathbb{E}[S|\mathcal{G}]$ should be a random variable, and it could be determined from the information in $\mathcal{G}$. Thus $\mathbb{E}[S|\mathcal{G}]$ should be $\mathcal{G}$-measurable.
- It should indeed be an estimate of S , and thus should give the same average as S over all the sets in $\mathcal{G}$. If $\mathcal{G}$ has many sets, which provide a fine resoltion of the uncertainty inherent in ω, then this partial averaging over the sets in $\mathcal{G}$ would mean that $\mathbb{E}[S|\mathcal{G}]$ is a good estimator of S.If $\mathcal{G}$ has only a few sets, then $\mathbb{E}[S|\mathcal{G}]$ is a crude esimate of S.

These required properties motivate the standard definition of Conditional Expectation, given in the next section.

III. Definition and Existence.

Definition 2. Let $(\Omega, \mathcal{F}, P)$ be a probability space, let $\mathcal{G}$ be a $sub - \sigma - algebra$ of $\mathcal{F}$, and let S be a random variable that is either nonnegative or integrable. The conditional expectation of S given $\mathcal{G}$, denoted by $\mathbb{E}[S|\mathcal{G}]$, is any random variable that satisfies :

1) (Measurability) $\mathbb{E}[S|\mathcal{G}]$ is $\mathcal{G}$-measurable, and
2) (Partial Averaging)

$$\int_C \mathbb{E}[S|\mathcal{G} \, dP = \int_C S \, dP \, for \, all \, C \, \in \, \mathcal{G}$$

Remark 3. This is by no means a constructive definition. Though function which satisfies the two properties will be called the conditional expectation, we have not shown the existence of such a function.

Lemma 4. *If f belongs to $\mathcal{L}$ and λ is defined on $\mathbb{X}$to $\Re$ by*

$$\lambda(E) = \int_E f d\mu,$$

then λ is a signed measure.[1] *If f belongs to $M^+(X, \mathbb{X}, \mu)$, then λ is a measure.*[2]

Theorem 5. *RADON-NIKODYM THEOREM. Let λ and μ be $\sigma - finite$ measures defined on $\mathbb{X}$ and suppose that λ is absolutely continuous with respect to μ. Then there exists a function f in $M^+(X, \mathbb{X}, \mu)$ such that $\lambda(E) = \int_E f d\mu \forall E \in \mathbb{X}$.*

Moreover, the function f is uniquely determined $\mu - almost$ everywhere.[3]

Proof of Existence of Conditional Existence by Radon-Nikodym Theorem

Theorem 6. *Let S be an extended random variable on $(\Omega, \mathcal{F}, P)$, $\mathcal{G}$ be a $sub - \sigma - algebra$ of $\mathcal{F}$. Assume that $\mathbb{E}(S)$ exists. Then there is a function, say Y_G : $(\Omega, \mathcal{G}) \rightarrow (\Re, \mathbb{B}(\Re))$, such that, for each $C \in \mathcal{G}$:*

$$\int_C Y_G dP = \int_C S dP$$

Any two such functions must concide a.e. $[\mathbb{P}]$. (Note : $Y_G$is $\mathcal{G}$-measurable)

[1] see Bartle for definition of a charge and lebesgue integral

[2] Robert G. Bartle,"The Elements of Integration and Lebesgue Measure", Pg 34, 42

[3] This can be extended to when $\lambda$is a signed measure. In this case f will belong to $M(X, \mathbb{X}, \mu)$

Proof: Define $\lambda(C)$ as :

$$\lambda(C) = \int_C SdP,\, C \in \mathcal{G} \tag{1}$$

Then λ is a set function on $\mathcal{G}$. Also, then λ is absolutely continuous with respect to P. (Lemma 4)

By Radon-Nikodym theorem (Theorem 4) ,$\exists$ a function, say $Y_G$in $M(\Omega, \mathcal{G}, P)$ such that :

$$\lambda(C) = \int_C Y_G dP \,\forall C \in \mathcal{G} \tag{2}$$

and which is unique a.e. w.r.t P.

By (1) and (2), we get $\int_C Y_G dP = \int_C SdP \,\forall C \in \mathcal{G}$ ■

IV. A Second Approach

Intuitively, our objective is to get as close as possible to the random variable S , but using only random variables that are measurable with respect to some smaller $\sigma - algebra$. We can use this to define conditional expectation in an alternative way :

Definition 7. Let $\mathcal{G} \subseteq \mathcal{F}$ be $\sigma - algebras$ and S be a random variable on $(\Omega, \mathcal{F}, P)$. Assume $< \infty$. Then the conditional expectation is an almost surely unique $\mathcal{G}$-measurable Y_G such that $\mathbb{E}[(S - Y_G)^2] = inf_z \mathbb{E}[(S - Z)^2]$, where the infimum is over all $\mathcal{G}$-measurable random variables. Note: We denote the minimizing Y_G by $\mathbb{E}[S|\mathcal{G}]$.

Remark 8. This is a weaker definition than the one before, since it assumes $\mathbb{E}(S^2) < \infty$.

Claim 9. The Y_G in Definition 7 satisfies, for each $C \in \mathcal{G}$, :

$$\int_C Y_G dP = \int_C SdP$$

Proof: We need $\int_C S - Y_G \, dP = 0$ or $\int (S - Y_G) I_C \, dP = 0$

or $\mathbb{E}[(S - Y_G) I_c] = 0$.

Consider the function $g(\lambda) = \mathbb{E}[(S - Y_G - \lambda I_c)^2]$

By definition of Y_G, this achieves a minimum at $\lambda = 0$. Hence $\frac{dg}{d\lambda}(0) = 0$.

=> $\mathbb{E}[(S - Y_G) I_c)] = 0$, as required. ■

Claim 10. If $\mathcal{G} = \{\Omega, \emptyset\}$, then $Y_G = \mathbb{E}(S)$

Proof: Only constant random variables are measurable w.r.t the trivial $\sigma - algebra$. Let $Z = c$. $\min_c \mathbb{E}[(S - c)^2] = \min_c \{var(S) + \mathbb{E}[S - c]^2\}$ which results in $c = \mathbb{E}(S)$. ■

Claim 11. If $\mathcal{G} = \{\Omega,\, A\, A^{\complement}\, \emptyset\}$, for some event A. then

$$Y_G = \begin{cases} \mathbb{E}[SI_A]/P(A) \; for \quad \omega \in A \\ \mathbb{E}[SI_B]/P(B) \; for \quad \omega \in B \end{cases}$$

Proof: In this case, only the random variables that take constant value a on A and b on B are $\mathcal{G}$-measurable.

$\mathbb{E}[(S - Z)^2] = \mathbb{E}[S^2 I_A] - 2a\mathbb{E}[S^2 I_A] + a^2 P(A) + \mathbb{E}[S^2 I_B] - 2bE[S^2 I_B] + b^2 P(A)$

Minimizing w.r.t a and w.r.t. b, we get the desired values.

Claim 12. If $\mathcal{G}$ is generated by a finite partition $\{A_1,\, A_2, \ldots A_n\}$ of the probability space Ω, then

$$Y_g = \sum_1^n c_i I_{A_i}(\omega)$$

$$where\, c_i = \mathbb{E}[SI_{A_i}]/P(A_i)$$

Proof: The proof is just the same as that of Claim 10, but in this case the random variable will take n values. ■

Remark 13. We have defined conditional expectation for random variables with finite variance. We can extend the definition to non-negative random variables, by choosing a sequence of simple random variables which converge to it from below. Simple random variables have finite variances, and hence a sequence of simple random variables will have a corresponding sequence of conditional expectation. By taking the supremum, of this sequence, we can define conditional expectation for non-negative random variables. Also, by representing a general integrable random variable as a difference of non-negative random variables, we can extend our definition to include these as well.

■

V. A Topological Approach

The above definitions are basically two approaches to the same concept. I will now generalise and combine the two approaches in a more abstract setting, and will show that the existence of Conditional Expectation is a direct consequence of the special structure of the function space $\mathcal{L}_2(\Omega, \mathcal{F}, P)$. We'll be needing the following results and definitions:

Theorem 14. *$If 1 \leq p < \infty$, then the space $\mathcal{L}_\mathcal{P}$is a complete normed linear space (Banach) under the norm*

$$\parallel S \parallel_p = \left\{ \int |S|^p \right\}^{\frac{1}{p}}$$

Proof: Robert G. Bartle.*The Elements of Integration and Lebesgue Measure*. Theorem 6.14. Pg 69. ■

Theorem 15. *Hilbert Space Projection Theorem. For every point x in a Hilbert space H and every closed convex set $C \subset H$, there exists a unique point z_0 for which $\parallel x - z \parallel$is minimized over C. This is, in particular, true for any closed subspace M of H. In that case, a necessary and sufficient condition for y is that the vector $x - y$ be orthogonal to M.*[4]

Lemma 16. *L_2 spaces are Hilbert Spaces with the inner product being defined by :*

$$< X, Y > = \int_C XY \, dP \, or \, \mathbb{E}[XY]$$

Proof: $< S, S >= \mathbb{E}[S^2] = \parallel S \parallel_2^2$ which is 0 only when [S]=0.

That the inner product defined above is linear in the first argument follows from the linearity of Lebesgue Integral.

Completeness follows from Theorem 13. ■

Lemma 17. *$\mathcal{L}_2(\Omega, \mathcal{G}, P)$ is a closed subspace of the function space $\mathcal{L}_2(\Omega, \mathcal{F}, P)$, and thus a Hilbert Space in itself.*

Proof: $\mathcal{L}_2(\Omega, \mathcal{G}, P)$ is complete. Since $\mathcal{G} \subseteq \mathcal{F}$, $\mathcal{L}_2(\Omega, \mathcal{G}, P)$ is a also a subspace of the function space $\mathcal{L}_2(\Omega, \mathcal{F}, P)$.Complete subspace => Closed subspace. (Take

[4]Look at appendix for related results

an accumulation point, use axiom of choice on the family of its neighbour hoods with radius 1/n , and we get a sequence which converges to itself. Use completeness) ■

A. Proof of Existence of Conditional Expectation based on Hilbert Space Projection

Theorem 18. *Take* $S \in$ *Hilbert Space* $\mathcal{L}_2(\Omega, \mathcal{F}, P)$, and $\mathcal{L}_2(\Omega, \mathcal{G}, P)$ as the closed subspace of this Hilbert Space. Hence, there exists a unique $Y_G \in \mathcal{L}_2(\Omega, \mathcal{G}, P)$ such that $\| S - Z \|_2$ is minimized over $\mathcal{L}_2(\Omega, \mathcal{G}, P)$.

Hence $\| S - Z \|^2$ is also minimized
=> $\int_\Omega (S-Z)^2 dP$ is minimized
=> $\mathbb{E}[(S-Z)^2]$ is minimized.
Moreover, $I_c \in \mathcal{L}_2(\Omega, \mathcal{G}, P) \; \forall \; C \in \mathcal{G}$
=> $S - Y_G \perp I_C \; \forall \; C \in \mathcal{G}$ (By Hilbert Space Projection theorem)
=>$< S - Y_G, I_C >= 0 \; \forall \; C \in \mathcal{G}$
=> $\int (S - Y_G).I_C \, dP = 0 \; \forall \; C \in \mathcal{G}$
=> $\int_c (S - Y_G) \, dP = 0 \; \forall \; C \in \mathcal{G}$
=> $\int_C S \, dP = \int_C Y_G \, dP \; \forall \; C \in \mathcal{G}$

Remark 19. We have assumed that $\mathbb{E}(S^2) < \infty$.But we can extend the proof and show the existence of conditional expectation in a manner similar to Remark 13. An important argument will be that if (ϕ_n) is an increasing sequence of simple random variables converging to S, then $\mathbb{E}[\phi_n|\mathcal{G}]$ will also be a.e increasing. To see this, for any n , let A be the set where $\mathbb{E}[\phi_{n+1}|\mathcal{G}] < \mathbb{E}[\phi_n|\mathcal{G}]$. Then :

$$\mathbb{E}[(\mathbb{E}[\phi_{n+1}|\mathcal{G}] - \mathbb{E}[\phi_n|\mathcal{G}])I_A] = \mathbb{E}[(X_{n+1} - X_n)I_A]$$

The L.H.S is less than or equal to zero, while the right hand side is greater than or equal to zero. Hence $P(A) = 0$.

VI. Conclusion

Though its existence is proved through Radon-Nikodym theorem in measure-theory literature, we have shown that the condition expectation of a random variable given a $\sigma\!-\!algerba$ arises naturally out of the structure of $\mathcal{L}_2$ Hilbert Spaces. It is nothing but the orthogonal projection of the random variable on the $sub-\sigma-algebra$. It is the *unique a.e best approximation* of the random variable on the closed subspace of the space of all random variables with finite variance.

Appendix

Theorem 20. *Let* X *be an inner product space and* $A \subseteq X$ *a complete, convex and non-empty subset. Then for every* $x \in X$ *, there exists a unique best approximation of* x *in* A *, i.e. there exists a unique element* $a_0 \in A$ *such that*

$$\| x - a_0 \| = d(x, A)$$

Theorem 21. *Let* X *be an inner product space, and* $A \subseteq X$ *a subspace and* $x \in X$ *. The following statements are equivalent:*

- $a_0 \in A$ is the best approximation of x in A .
- $a_0 \in A$ and $x - a_0 \perp A$. Thus, the best approximation of x in a subspace A is just the orthogonal projection of x in A .

Theorem 22. *Let* X *be an Hilbert space and* $A \subseteq X$ *a closed subspace. Then the orthogonal complement of* A *, denoted* $A^\perp$ *, is a topological complement of* A *. That means* A *is closed and* $X = A \oplus A^\perp$.

References

[1] Robert G. Bartle.*The Elements of Integration and Lebesgue Measure* (Wiley Classics Library Edition 1995)
[2] S.E Shreve. *Stochastic Calculus for Finance - Volume 2*. Chapters 1-2. (Springer, 2004)
[3] Robert B. Ash and Catherine A.Doleans-Dade, Probability and Measure Theory.(Elsevier)
[4] http://planetmath.org