oeruhnaornf

Monroe Stephenson

November 2024

1 conditional expectation

conditional expectation, denoted as $E[X|G]$ is another random variable that averages out X given the information in G More formally,

$$E[X|G] = \int_A E[X|G] = \int_A X dP \text{for all } A \in G$$

This is taking the average of expected value of X

Example 1. *Suppose X is the result of a rice role, and G is the information that tells us wheter the roll was even or odd, then $E[X|G]$ is the average of X given whether it was an odd or even roll. So if even,*

$$E[X|G] = 4 \textit{if even } E[X|G] = 3 \textit{ if odd}$$

2 Monotone class theorem

Theorem 1. *Suppose M is a monotone class, i.e. a collection closed under two types of limit operations, countable increading or countable decreasing. In other words, if $A_1, \ldots, \in M$, is a sequence of sets in M such that $A_1 \subset A_2 \subset ...$, then the union $\bigcup_{n=1}^{\infty} A_n$ is also in M Similarly, if $B_1, \ldots \in M$, such that $B_1 \supset B_2 \supset \ldots$, then the intersection, $\bigcap_{n=1}^{\infty} B_n$ is also in M. Then, if M contains an algebra A (closed under finite unions, intersections, and complements), then M must contain the entire σ-algebra generated by A. In other words, if you start with a collection of simple sets, an algebra, and build up to more complex sets by taking countable unions and intersections, then the monotone class will contain all of these resulting sets.*

This is usual because we can prove things on a simple set, like a property of something, on an algebra, then extend it to σ-algebras far more easily. Like using this to establish theorems in conditional expectations ny proving them first for simple events, then extending it to the full σ algebra

3 Regular Conditional Probability

We want to understand the Probability of an random variable Y given X more often than not, and for discrete random variables we can easily define conditional probabilities such as $P(Y = y|X = x)$ by dividing probabilities. However with continuous random variables or more complex spaces, defining $P(Y \in A|X = x)$ in a meaningful, measurable, way is challening. This is because the probability of a continuous random variable taking any single value is 0, $P(X = x) = 0$ for a continuous X, making the definitions impossible. A regular conditional probability is a function that allows us to condition on the event $X = x$ even in continuous settings by definining conditional probabilities in a way that is measurable and consistent with the properties of a probability measure.

Definition 1. *Let (X, Y) be a pair of random variables on a Probability space, (Ω, F, P), and let G be the σ-algebra generated by X, so $G = \sigma(X)$. A regular conditional probability, $P(Y \in A|X = x)$, is a function that satisfies the following properties.*

1. *Probability measure: For each fixed value x, $A \mapsto P(Y \in A|X = x)$ is a probability measure on the space where Y is defined.*

2. *Measurability: For each set A in the range of Y, the function $x \mapsto P(Y \in A|X = x)$ is measurable with respect to the σ-algebra generated by X.*

3. *Consistency: For any event A in the range Y, the regular conditional probability should satisfy the law of total probability,*

$$P(Y \in X) = \int_\Omega P(Y \in A|X = x) dP_X(x)$$

   where P_X is the probability distribution of X.

This matters, because we can define conditional distributions rigourously when traditional conditioning doesnt work.
Note, the Markov property, they regular condtional probabilities help formalize the idea of memorylessnes. For example, in Markov chains, the conditional distribution of future states depends only on the present state, not the past states. Under certain conditions, like standard Borel spaces, then the regular conditional probabilities are guaranteed to exist. However, not all probability spaces admits regular conditional probability.

3.1 Construction of Regular condtional probabilities

Consider the discrete case of X, then we can just do

$$P(Y = y|X = X) = \frac{P(X = x, Y = y)}{P(X = x)}$$

However, for continuous the story is a different story. If X and Y have a joint density $f_{X,Y}(x,y)$ and a marginal density $f_X(x)$, then we have

$$P(Y \in A|X = x) = \int_A \frac{f_{X,Y}(x,y)}{f_X(x)} dy$$

which uses the Randon-Nikodym derivative and is the foundation for conditional densities.

3.2 examples of the regular conditional probabilities

Let X and Y be independent continuous random variables with respective densities $f_X(x)$ and $f_Y(y)$, In this case, fro any set A,

$$P(Y \in A|X = x) = P(Y \in A) = \int_A f_Y(y)dy$$

since X and Y are independent, knowing $X = x$ gives no additional information.

3.2.1 Dependent continuous variables

Suppose X and $Y = X + Z$, where Z is an independent, standard, normal variable. The regular conditional probability $P(Y \in A|X = x)$, would depend on X.

$$P(Y \in A|X = x) = P(X + Z \in A|X = x) = P(Z \in A - x)$$

where $A - x = \{y - x \mid y \in A\}$. In other words, once we know that $X = x$, we only need to consider the distribution of X, shifted by x, to determmine Y's distribution condtioned on $X = x$.

3.3 Bayesian Statistics

Regular conditional probability is crucial in Bayesian inference, wehre we need to update probabilities based on new information. Given a prior distribution, $P(\theta)$ for a parameter θ and a likelihood function $P(Y|\theta)$, the posterior distribution $P(\theta|Y = y)$ is a regular conditional probability that allows us to formally condition on observed data, even when both θ and Y are continuous.

4 Conditional expectation more in depth:

Conditional expectation of a random variable given a sigma algebra G, (which is a set of known unformation.), denoted by $E[X|G]$, which is a best guess of X given the information.

Definition 2. *For a random variable $X \in L^1$, so the expectation of X is finite, and a subalgebra $G \subset F$, the conditional expectation $E[X|G]$ is the unique G measurable function Y that satisfies $E[Y 1_A] = E[X 1_A]$ for all $A \in G$ so $E[X|G]$ is the best fit of X to the information in G in the sense that it matches the average value of X over all sets in G*

We think of G as containing information about events that have occured and $E[X|G]$ as the revised mean of X given the infomation. You can imagine G represents observations up to a certain time in a stochastic process, then $E[X|G]$ is the best prediction of X given what we've seen so far. For example, if X is tommorows stock price, and G represents todays market info, then $E[X|G]$ is the expected stock price based on todays data.

4.1 Key properties of conditional expectations

1. **Linearity**: If X and Y are integrable random variables, and a, b are constants, then

$$E[aX + bY|G] = aE[X|G] + bE[Y|G]$$

2. **Taking out what is known**: if Z is G-measurable, then

$$E[ZX|G] = Z \cdot E[X|G].$$

   this is because Z can be taken out, since it is fully determined by G.

3. **Law of Interated Expectations (Tower Property)**: If $H \subset G$ are two nested $\sigma-$algebras, then

$$E[E[X|G]|H] = E[X|H]$$

   this property allows us to compute expectations in a step by step manner.

4. **Jensen's Inequality**: If u is a convex function, and $X \in L^1$, then $E[u(X)|G] \leq u(E[X|G])$. This allows us to bound expectations and shows that applying a convex function outside of the expectation gives a larger value than applying it inside.

4.2 Examples

Consider a rv X representing the outcome of rolling a fair 6 sided die, with each probability being $\frac{1}{6}$, then let G be the sigma algebra telling us if X is even or odd, then

$$E[X|X \text{ is even}] == \frac{2+4+6}{3} = 4$$
$$E[X|X \text{ is odd}] == \frac{1+3+5}{3} = 3$$

Consider a random variable $Y \in L^2$, and let K be a closed subspace of L^2 generated by the set of G-measurable functions. The conditional expectation $E[Y|G]$ is the othogonal project of Y onto K. The projection property hels us find the best approximation of Y in terns of the information in G. Martingales are closely related to conditional expectation. A sequence of random variables, $\{X_n\}$ is a martingale with respect to a filtration $\{F_n\}$ is

$$E[X_{n+1}|F_n] = X_n.$$

This property means that the best estimate of X_{n+1} given all of the information up to time n is just X_n.

5 Regular Conditional Probability

In this section we fix a probability space (Ω, F, P) and a sub sigma field, $G \subseteq F$, and a random variable $X\colon (\Omega, F) \to (E, \epsilon)$ taking values in some measure space (E, ϵ). Our goal is to define and prove the existence of regular conditional probabilities.

As a motivation let us fix some $B \in \epsilon$, then $1_B(X)$ is a bounded random variable. Thus, the conditional expectation $E[1_B(X)|G]$ exists and is $[0,1]$-valued almost surely. We define,

$$P(X \in B|G) := E(1_B(X)|G).$$

Now let us consider a sequence of disjoint sets $(B_n) \subset \epsilon$. By the linearity and monotonicity of conditional expectations,

$$P\left(X \in \bigcup_n B_n|G\right) = \sum_n P(X \in B_n|G)$$

almost surely. Unfortunately, this identity only hold outside of a set N of measure zero that depends on the sequence (B_n). Likewise, for any $B \subset \epsilon$, there exists a set N_B of measure zero such that

$$P(X \in B|G)(\omega) \in [0,1] \text{ for all } \omega \in \Omega \setminus N_B.$$

Our question is now if we choose versions of the conditional expectations $P(X \in B|G)$, $B \in \epsilon$, in a consistent way such that

$$B \mapsto P(X \in B|G)(\omega)$$

is a probability for almost all $\omega in \Omega$. The answer will in general be negative because we have uncountable many of the sets N_B and the union of uncountably many sets of measure zero sets may not have measure zero. However, we can do this if (E, ϵ) is a nice enough space.

Definition 3. *Let (Ω_i, F_i) $i = 1, 2$ be measurable spaces. A mapping $\kappa : \Omega_1 \times F_2 \to [0,1]$ is called a stochastic kernel from (Ω_1, F_1) to (Ω_2, F_2) if the following two properties hold.*

1. *For any $B_2 \in F_2$, the mapping $\omega \mapsto \kappa(\omega, B_2)$ is F_1-measurable.*
2. *For any $\omega \in \Omega_1$, the mapping $B \mapsto \kappa(\omega, B)$ is a probability measure on (Ω_2, F_2)*

So we are going between two measurablilitu spaces. The stochastic kernek is a function that assigns probabilities to events in F_2 based on values in Ω_1.

The two properties, have specific meaning. The first, Measurability, means that for every event B_2, $\kappa(\cdot, B_2)$ behaves like a measurable function from Ω_1 to the interval $[0, 1]$, ensuring it is compatible with the structure of (Ω_1, F_1). This ensures κ is measurable as a function of ω, allowing us to use κ in integration and other operations involving F_1. The second, is that for every given fixed point $\omega \in \Omega_1$, the mapping $B \mapsto \kappa(\omega, B)$ is a probability measure on (Ω_2, F_2). So for every ω, the function $\kappa(\omega, \cdot)$, assigns probabilities to events in F_2.

Together, these properties mean that κ behaves like a conditional probability function, specifying the probability in F_2 as a function of points in Ω_1.

Intuition, the stochastic kernel can be thought of as a conditional probability. For a fixed event $B \in F_2$, then $\kappa(\omega, B)$ gives the probability that an outcome falls in B depending on a point $\omega \in \Omega_1$. This is commonly ised to model the probability of tranditioning from one space to events in another space where you start.

In the context of Markov chains, a stochastic kernel can represent transition probabilities, where Ω_1 represents the current state, and Ω_2 represents the possibile next states.

It is often sufficent to instead check of Measurability on all B, we can verify this property for a $\cap$-stable generator F_2.

A $\cap$-stable generator is a collection of sets within F_2 that is closed under intersections and can generate the entire F_2 through unions, intersections, and complements.

Since measurable functions and measures defined on generators can often be extended to the full $\sigma - algebra$, proving Measurability for a generator suffices to ensure it holds for all sets in F_2. This approach simplifies proving Measurability by focuing on smaller more manageable collection of sets rather than F_2.

Example 2. *Consider a markov process where Ω_1 represents the current state of a system, say the state of a particle at a time t. Now consider Ω_2 represents possbile future states, i.e. the state of the particle at time $t + 1$.*

The stochastic kernel $\kappa(\omega, B)$ would then give the probability of transitioning from the current state $\omega \in \Omega_1$ to a set of states $B \subseteq \Omega_2$ in the nect step. this transtion probabilities depends on ω and allows us to model how the system evolves over time.

Proof. Now we prove that it suffices to the measurability of $\cap-$stable generator.

We want to show that every mapping $\omega \mapsto \kappa(\omega, B_2)$ is F_1-measurable for every $B_2 \in F_2$. However, instead of verifying measurability for each B_2 in the possibly large $\sigma-$algebra F_2, we use a smaller collection of sets, specifically, $\cap$-stable generator of F_2.

Note, a probability space is complete if every subset of any null set (a set of measure zero) is also measurable. This also ensures that almost surely limits of sequences of measurable functions are also measurable, whcih is crucial in our proof.

Let C be a collection of all sets $B_2 \in F_2$ for which the mapping $\omega \mapsto \kappa(w, B_2)$ is F_1-measurable.

Now we need to show that C is a monotone class, i.e. closed under countable increasing and decreasing limits.

countable increasing limits, if $B_2^n \subset B_2^{n+1}$ for all n, and $B_2 = \bigcup_{n=1}^{\infty} B_2^n$, then $\kappa(\omega, B_2) = \lim_{n\to\infty} \kappa(\omega, B_2^n)$. Since each $\kappa(\omega, B_2^n)$ is measurable (by the assumption that $B_2^n \in C$), the limit will also be measurable.

Similarly, if $B_2^n \supset B_2^{n+1}$ and $B_2 = \bigcap_{n=1}^{\infty} B_2^n$, then $\kappa(\omega, B_2) = \lim_{n\to\infty} \kappa(\omega, B_2^n)$. By completeleness of the probability space, this almost surely limit is measurable.

Let $G \subset F_2$ be a $\cap-$stable generator of F_2. For each $B_2 \in G$, we assume or verify that $\kappa(\omega, B_2)$ is F_1-measurable. (Our assumption here)

Since C is a monotone class, and contains G_1, a $\cap-$stable generator of F_2, the monotone class theorem imples that C must contain the σ-algebra F_2. We conclude that $\omega \mapsto \kappa(\omega, B_2)$ is F_1-measirable for all $B_2 \in F_2$.

Note, G is the $\cap$-stable generator for F_2, so then $\sigma(G) = F_2$. C is the monotone class we define for the prupose of applying the monotone class theorem. C is the set of all sets B_2 for which the function is measurable. By definition, C includes precisely those sets for which we have already verified the desired measurability property.

In short, we begin by assuming that $G \subset C$, assuming $\kappa(\omega, B_2)$ is measurable for all $B_2 \in G$, then by showing C is a monotone class, we can apply the monotone class theorem to conclude that C contains all of F_2. Thus, we can extend measurability from G to all of F_2.

The completeness of the probability space is crucial. Without it, the almost sure limit of measurable functions might fail to be measurable, which would mean that C may not be closed under limits.

completeleness ensures that any set of measure zero, where Measurability might fail, has all its subsets included in F_1, maintaining measurability of of as limits and allowing us to use the monotone class approach. $\square$