Course : Mathematical Statistics
Lecture 2: Sufficient Statistics and Information Inequality

Contents

1 Introduction

2 Radon Nikodym Theorem

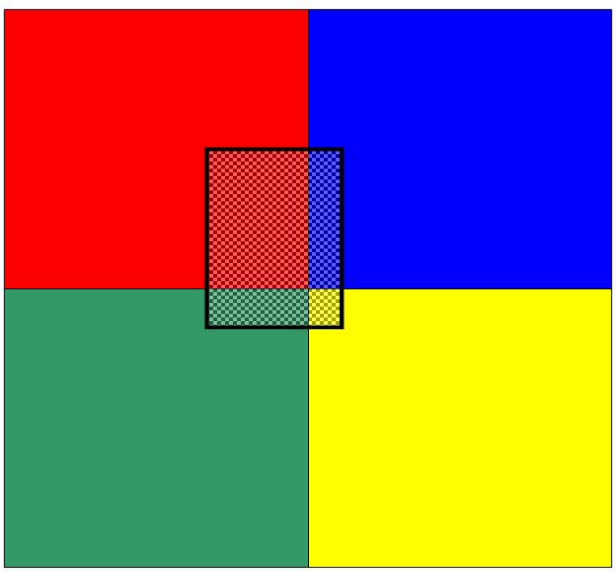
3 Conditional Expectation

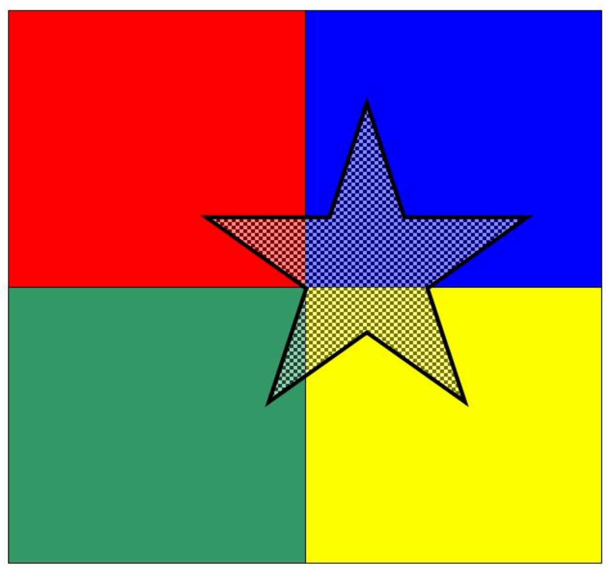
Conditional expectation had been defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

in terms of the probability of the events. It was Kolmogrov, in his Foundations of Theory of Probability, gave the interpretation of conditional probabilities as random variables, based on the axiomatic approach to probability. Let us understand the concept of conditional probability with the help of the following example. Consider a slightly modified dart game with the following modification:

- the dart board is square and has a checkered region as shown in the figure below.The board is divided into 4 regions, indicated by the 4 colours.
- The player should throw the dart blind-folded
- The player should hit the dart in the checkered region to win. The score is based on the distance from the checkered region. Closer the dart is to the checkered region, higher will be the score.
- One game consists of 3 chances to hit the board.
- At the begining of each game, the shape of the checkered region is changed and the player sees the region and is then blinded folded.
- When the player throws the dart, there is a game arbitrator who tell s the colour of the region where the dart hit. This acts a guidance to the player in his next throw.



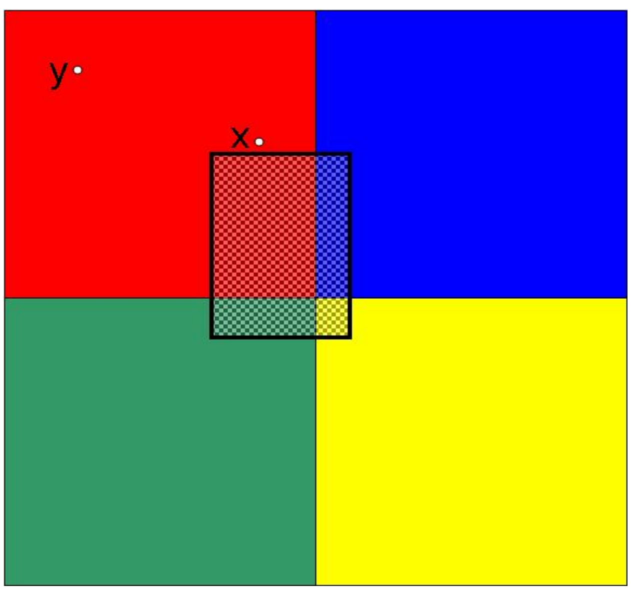


Let us analyse this game from the mathematical point of view. The square board forms the outcome of each throw, in other words, the sample space Ω. Let $\{\Omega, \mathcal{F}, P\}$ be the measure space associated with probability measure P. Let us name the four disjoint coloured regions, A_R, A_B, A_G, A_Y. These four regions are such that $\Omega = A_R \cup A_B \cup A_G \cup A_Y$. Let us consider the sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$, formed by the finite union of the sets A_R, A_B, A_G, A_Y.
Notice the following:

- The σ-algebra $\mathcal{G}$ consists only of those sets (events) about which the arbitrator can give information to the player. For example,"Did the dart hit red?","Did the dart hit yellow?", "Did the dart hit green", "Did the dart hit blue?", "Did the dart hit yellow or red?" ,"Did the dart hit top half of the board?", this question is answer if the arbitrator calls out the colour, the player would know whether the dart hit top half or bottom half. Essentially, $\mathcal{G}$ models the information that the player will receive from the arbitrator. This shows how σ-algebra models the information, which justifies the statement that we often come across in statistics and probability that the information is contained in the σ-algebra.

- Notice that the arbitrator does not give any information about what is the distance of the current hit from the checkered region. All the arbitrator does is call out the colour of the region where the dart has hit.

So the player proceeds is, at the begining of the games, he sees the shape of the checkered region and calculates beforehand the probability of winning associated with each colour. Let the checkered region be denoted by C. Even though the arbitrator does not give the position of the dart, but just the colour, the best choice for the player is to throw the dart into the region with the highest probability of winning. With each point on the board, the player can associate what is the probability of winning better based on arbitrator's information, if the dart hits that particular point.



Even though a point x is closer to the checkered region than point y, since the player does not get that proximity information from the arbitrator, for player hitting anypoint in red is equivalent and so he associates the same probability of winning to any point in red. So to each point $\omega \in \Omega$, the player associated a value as given below:

$$f(\omega) = \begin{cases} P(C|A_R) \text{ if } \omega \in A_R \\ P(C|A_Y) \text{ if } \omega \in A_Y \\ P(C|A_B) \text{ if } \omega \in A_B \\ P(C|A_G) \text{ if } \omega \in A_G \end{cases} \tag{1}$$

This $f(\omega)$ is the conditional probability of C given $\mathcal{G}$.

3.1 General Case

In the general case, we do not expect that the σ-algebra $\mathcal{G}$ comes from such a partition. But what is important here is to understand that $\mathcal{G}$ models the information that is available. Now instead of defining the conditional probability based on the partitions of Ω, we define a new measure ν. To do so let us fix a set $C \in \mathcal{F}$ and define a finite measure ν on $\mathcal{G}$.

$$\nu(G) = P(C \cap G) \text{ for all } G \in \mathcal{G} \tag{2}$$

Radon Nikodym..
Definition
Gambling Interpretation

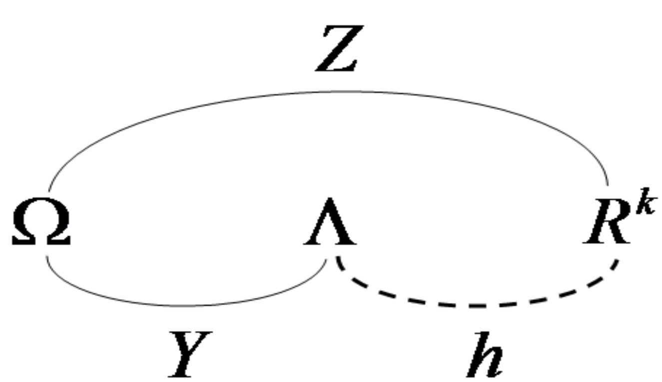
Let $\mathcal{C}$ be a collection of subsets of Λ. We define

$$f^{-1}(\mathcal{C}) = \{f^{-1}(C) : C \in \mathcal{C}\} \tag{3}$$

Definition (Measurable function): Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be measurable spaces and f a function from Ω to Λ. The function f is called a measurable function from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ if and only if $f^{-1}(\mathcal{G}) \subset \mathcal{F}$.

If f is measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$, then $f^{-1}(\mathcal{G})$ is a sub-σ-field of $\mathcal{F}$. It is called the σ-field generated by f and is denoted by $\sigma(f)$.

Lemma 1.2: let Y be a measurable from $(\Omega, \mathcal{F})$ to $(\Lambda, \mathcal{G})$ and Z a function from $(\Omega, \mathcal{F})$ to $\mathcal{R}^k$. Then Z is measurable from$(\Omega, \sigma(Y))$ to $(\mathcal{R}^k, \mathcal{B}^k)$ if and only if there is a measurable function h from $(\Lambda, \mathcal{G})$ to $(\mathcal{R}^k, \mathcal{B}^k)$ such that $Z = h \bigcirc Y$.
Proof: Given that $Y : \Omega \to \Lambda$ is measurable from $\langle \mathcal{F}, \mathcal{G} \rangle$.
Only If: If Z is measurable w.r.t $\langle \mathcal{F}, \mathcal{R}^k \rangle$, then show that there exists a function h such that $Z = h \bigcirc Y$ and h is measurable w.r.t $\langle \mathcal{G}, \mathcal{R}^k \rangle$.
Define h such that $h(\nu) = Z(\omega)$, if $\nu = Y(\omega)$, $\nu \in \Lambda$. Now we have to show that h is measurable w.r.t $\langle \mathcal{G}, \mathcal{R}^k \rangle$. That is, we have to show that, $h^{-1}(B) \in \mathcal{G}$ for all $B \in \mathcal{B}^k$.



4 Sufficiency and Minimal Sufficiency

Let us consider the the dart game example in the previous section. Suppose we are trying to model the players in terms of whether they have a bias towards hitting the dart to the top half of the board or bottom half. Now the arbitrator calls out the colour of the box where the dart hits the board.

4.1 Sufficient Statistic

4.2 Minimal Sufficiency

4.3 Dominated Families

In the next section we will be looking at the factorization theorem, which will give us a way to obtain the sufficient statistic instead of guessing and then finding out the whether the conditional probability is independent of the θ or not. This requires that the density function of the measure be available. At this point we get a question, whether every probability measure has a density function with respect to some σ-finite measure.(Can't believe, that I got this question a few days back and this question is actually being answered).
Definition: Let $\mathcal{M}$ be a family of measures defined over a measurable space $(\Omega, \mathcal{F})$. Then $\mathcal{M}$ is said to be *dominated* by a σ-finite measure μ defined over $(\Omega, \mathcal{F})$ if each member of $\mathcal{M}$ is absolutely continuous with respect to μ. The family $\mathcal{M}$ is said to be dominated if there exists a σ-finite measure dominating it.
The below theorem answers the above question as to when does a density function exist, that is when is a measure dominated.
Theorem: A family $\mathcal{P}$ of probability measures over a Euclidean space $(\Omega, \mathcal{F})$ is dominated if and only if it is separable with respect to the metric

$$d(P, Q) = \sup_{A \in \mathcal{F}} |P(A) - Q(A)|$$

or equivalently with respect to the convergence definition

$$P_n \to P \text{ if } P_n(A) \to P(A) \text{ uniformly for } A \in \mathcal{F}.$$

The above theorem can be extended from measures on Euclidean spaces to abstract spaces. The property that was essential in proving the above theorem was that $\mathcal{F}$ had countable generators, that is, for which there exists a countable number of sets B_i such that $\mathcal{F}$ is the smallest σ-field containing the B_i. So essentially the above theorem holds for any σ-field having this property. If $\mathcal{F}$ does not possess a countable number of generators, a somewhat weaker conclusion can be asserted.
Two families of measures$\mathcal{M}$ and $\mathcal{N}$ are equivalent if $\mu(A) = 0$ for all $\mu \in \mathcal{M}$ implies $\nu(A) = 0$ for all $\nu \in N$ and vice versa.

Theorem: A family $\mathcal{P}$ of probability measures is dominated by a σ-finite measure if and only if $\mathcal{P}$ has a countable equivalent subset.
Proof: Let us prove it for the finite case.

If part: Consider the probability measures defined on $(X, \mathcal{M})$. Let there be a countable equivalent subset of $\mathcal{P}$, say, $\{P_1, P_2, \cdots\}$. We need to show that there exists a measure λ such that $\mathcal{P}$ is dominated by λ.
Define λ to be

$$\lambda(E) = \sum_{i=1}^{\infty} \frac{P_i(E)}{2^i} \quad \text{for } E \in \mathcal{M} \tag{4}$$

$\mathcal{P}$ is dominated by λ.
Only If part: Consider the family $\mathcal{P}$ dominated by a finite measure λ. We need to show that there exists a family $\mathcal{Q}$ of measures such that $\mathcal{Q}$ is countable and $\mathcal{Q} \subset \mathcal{P}$.
The family $\mathcal{P}$ is dominated by λ, Radon Nikodym theorem implies that for each $\mu \in \mathcal{P}$ there exists a derivative $f_\mu = \frac{d\mu}{d\lambda}$. Let $K_\mu = \{x \in X : f_\mu(x) > 0\}$.
Consider the class $\mathcal{C}$ is sets of $\mathcal{M}$ such that

$$\mathcal{C} = \Big\{ C : C \in \mathcal{M} \text{ and there exists a measure } \nu \in \mathcal{P} \text{ such that } \nu(C) > 0 \text{ and } f_\mu > 0 \text{ a.e.} \lambda \text{ on } C \Big\} \tag{5}$$

Consider the supremum of the λ in $\mathcal{C}$.

$$m = \sup_{C \in \mathcal{C}} \lambda(C) \tag{6}$$

Consider a sequence of sets $\{C_i\}$, where $C_i \in \mathcal{C}$, such that, $\lim_{i\to\infty} \lambda(C_i) = m$.
Let $C_0 = \cup_{i=1}^{\infty} C_i$. We can verify that $C_0 \in \mathcal{C}$.
Let $\mu_i \in \mathcal{P}$ be the measures corresponding to C_i. Let $\mathcal{Q} = \{\mu_1, \mu_2, \cdots\}$. $\mathcal{Q}$ is a countable subset of $\mathcal{P}$. We need to show that $\mathcal{Q}$ is equivalent to $\mathcal{P}$, that is, $\mathcal{Q} << \mathcal{P}$ and $\mathcal{P} << \mathcal{Q}$.
To show that $\mathcal{Q} << \mathcal{P}$ is trivial. We need to show that $\mathcal{P} << \mathcal{Q}$.
Consider a set $E \in \mathcal{M}$ such that $\mu_i(E) = 0$ for all $\mu_i \in \mathcal{Q}$. We need to show that for any $\mu \in \mathcal{P}$, $\mu(E) = 0$. If $E \notin \mathcal{C}$, then the statement holds anyway because for a set which is not in $\mathcal{C}$, $\mu(E) = 0$ for all $\mu \in \mathcal{P}$. So let us consider $E \in \mathcal{C}$.

4.4 Factorization Theorem

It is inconvenient to have to compute the conditional distribution of X given the statistic in order to determine whether or not T is sufficient. A simple check is provided by the following factorization criterion. Consider first the case that X is discrete, and let $P_\theta(x) = P_\theta\{X = x\}$. Then a necessary and sufficient condition for T to be sufficient for θ is that there exists a factorization

$$P_\theta(x) = g_\theta[T(x)]h(x) \tag{7}$$

where the first factor may depend on θ but depends on x only through $T(x)$, while the second factor is independent of θ.

The factorization criterion of sufficiency, derived above for the discrete case, can be extended to any dominated family of distributions, that is, any family $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$ possessing probability densities p_θ with respect to some σ-finite measure μ over $(X, \mathcal{A})$. The proof of this statement is based on the existence of a probability distribution $\lambda = \sum_i c_i P_{\theta_i}$, which is equivalent to $\mathcal{P}$ in the sense that for any $A \in \mathcal{A}$

$$\lambda(A) = 0 \text{ if and only if } P_\theta = 0 \text{ for all } \theta \in \Omega.$$

5 Complete Statistic

5.1 Basu's Theorem

6 UMVUE

7 Information Inequality and Cramer Rao Bound

8 Problems

8.1 Sufficient, Minimal, Complete Statistics

1. **(Factorization Theorem and Order Statistics)** Let $X_1, \cdots, X_n$ be i.i.d. random variables having a distribution $P \in \mathcal{P}$, where P is the family of distributions on $\mathbb{R}$ having continuous c.d.f.s. Let $T = (X_{(1)}, \cdots, X_{(n)})$ be the vector of order statistics. Show that, given T , the conditional distribution of $X = (X_1, \cdots, X_n)$ is a discrete distribution putting probability $1/n!$ on each of the $n!$ points $(X_{i_1}, \cdots, X_{i_n}) \in \mathbb{R}^n$, where $\{i_1, \cdots, i_n\}$ is a permutation of $\{1, \cdots, n\}$; hence, T is sufficient for $P \in \mathcal{P}$.
   Note: The order statistics can be shown to be sufficient even when $\mathcal{P}$ is not dominated by any σ-finite measure but then the factorization theorem is not applicable.

2. **(Sufficient Statistics)** Show that if T is sufficient statistic and $T = \psi(S)$, where ψ is measurable and S is another statistic, then S is sufficient.

3. **(Complete and sufficient statistic is minimal)** Let T be complete (or boundedly complete) and sufficient statistic. Suppose that there is a minimal sufficient statistic S. Show that T is minimal sufficient and S is complete (or boundedly complete).

4. **(Relation between two statistics)** Let T and S be two statistics such that $S = \psi(T)$ for a measurable ψ. Show that

   (a) if T is complete, then S is complete;

   (b) if T is complete and sufficient and ψ is one-to-one, then S is complete and sufficient.

   (c) the results in (a) and (b) hold even if the completeness is replaced by bounded completeness.

5. **(Boundedly complete but not complete)** exercise 53

6. **(Complete but not sufficient)** exercise 45

7. **(Sufficient Statistics for derived families of distributions)** exercise 46

8. Let $X_1, \cdots, X_n$ be i.i.d from the $\mathcal{N}(\theta, \theta^2)$ distribution, where $\theta > 0$ is a parameter. Find a minimal sufficient statistic for θ and show whether it is complete.

9. Exercise 54

10. Exercise 44

11. **(Example 2.13, Exercise 2.47, 2.32)** Let $X_1, \cdots, X_n$ be i.i.d. random variables from P_θ, the uniform distribution $U(\theta, \theta+1), \theta \in \mathbb{R}$. Suppose that $n > 1$.

    (a) Show that $T = (X_{(1)}, X_{(n)})$ is sufficient for θ.

    (b) Show that T is minimal sufficient for θ.

    (c) Show that T is not complete.

    Note: A complete and sufficient statistic is minimal but a minimal statistic need not be complete.

12. exercise 51,29

13. exercise 28

14. exercise 30

15. exercise 43

16. exercise 28

8.2 Admissibility

1. **(Nonexistence of unbiased estimator)** exercise 84

2. exercise 82