Find $f(X)$ that minimizes $E[(Y - f(X))^2|X]$

Asked 3 years ago Active 3 years ago Viewed 2k times

2

1

> Let X and Y random variables with $E(Y) = \mu$ and $E(Y^2) < \infty$. Deduce that the random variable $f(X)$ that minimizes $E[(Y - f(X))^2|X]$ is $f(X) = E[Y|X]$.

I just find the minimum with derivatives

$$\frac{d}{df(X)}E[(Y - f(X))^2|X] = -2E[Y - f(X)|X]$$

$$= -2E[Y|X] + 2E[f(X)|X] = 0$$

$$\Leftrightarrow E[Y|X] = E[f(X)|X]$$

$$\Leftrightarrow f(X) = E[Y|X]$$

Is this right?

I founded this solution

$$\mathbb{E}[(Y-c)^2] = \mathbb{E}[Y^2 - 2Yc + c^2] = \mathbb{E}[Y^2] - 2c\mathbb{E}[Y] + c^2$$
$$= \mathbb{E}[Y^2] - 2c\mu + c^2.$$

Find the extreme point by differentiating,

$$\frac{d}{dc}(\mathbb{E}[Y^2] - 2c\mu + c^2) = -2\mu + 2c = 0 \quad \Rightarrow c = \mu.$$

Since, $\frac{d^2}{dc^2}(\mathbb{E}[Y^2] - 2c\mu + c^2) = 2 > 0$ this is a min-point.
b) We have

$$\mathbb{E}[(Y - f(X))^2 \mid X] = \mathbb{E}[Y^2 - 2Yf(X) + f^2(X) \mid X]$$
$$= \mathbb{E}[Y^2 \mid X] - 2f(X)\mathbb{E}[Y \mid X] + f^2(X),$$

which is minimized by $f(X) = \mathbb{E}[Y \mid X]$ (take $c = f(X)$ and $\mu = \mathbb{E}[Y \mid X]$ in a).
c) We have

$$\mathbb{E}[(Y - f(X))^2] = \mathbb{E}\big[\mathbb{E}[(Y - f(X))^2 \mid X]\big],$$

so the result follows from b).

Is this wrong too?

probability statistics derivatives expectation

edited Sep 5 '16 at 14:35

asked Sep 5 '16 at 13:23

Roland
1,380 1 12 34

13:29

@Did I wouldn't call it so wrong, nor innovative. It makes sense if one minimizes the square for a given (fixed) value of X - in that case it's the direct generalization of showing that the value of a that minimizes $E([X-a]^2)$ is $a=E(X)$. – leonbloy Sep 5 '16 at 13:37

@leonbloy What is $\frac{\partial}{\partial f(X)}$ already? You see. It seems that, despite your good heart, I will stick to "very far from being right"... – Did Sep 5 '16 at 13:39

See eg ocw.mit.edu/courses/electrical-engineering-and-computer-science/... – leonbloy Sep 5 '16 at 13:39

2 Answers

2

Let us try to make your argument more rigorously legitimate. First, let's agree that the goal is to find some (measurable) function $f:\mathbb{R}\to\mathbb{R}$ such that for any (measurable) function $g:\mathbb{R}\to\mathbb{R}$ the inequality

$$\mathbb{E}\left[(Y-f(X))^2 \mid X\right] \leq \mathbb{E}\left[(Y-g(X))^2 \mid X\right] \tag{1}$$

holds almost surely. Now, let Ω be the sample space on which X is defined and let $h(X):\Omega\to\mathbb{R}$ be a representative of $\mathbb{E}\left[Y \mid X\right]$ and $k(X)$ be a representative of $\mathbb{E}\left[Y^2 \mid X\right]$, each defined for every $\omega\in\Omega$. Then,

$$\mathbb{E}\left[(Y-f(X))^2 \mid X\right] = k(X) - 2f(X)h(X) + f(X)^2$$

(where equality means the RHS is in the equivalence class of the LHS). Now, for a fixed $\omega\in\Omega$, if we minimize

$$k(X)(\omega) - 2\lambda h(X)(\omega) + \lambda^2$$

in the variable λ, using differentiation as you have above, you will find that $\lambda = h(X)(\omega)$. Thus, for each $\omega\in\Omega$, defining $\lambda(\omega)=h(X)(\omega)$ minimizes the previous expression pointwise in Ω. Thus with $f(X)=h(X)$, Eq. (1) is minimized almost surely.

edited Sep 5 '16 at 14:11 answered Sep 5 '16 at 14:03

Tom

1

Hint: the typical trick is the minus-and-add trick: with $Z = E(Y|X)$,

$$\begin{aligned} E[(Y-f(X))^2|X] &= E[(Y-Z+Z-f(X))^2|X] \\ &= E[(Y-Z)^2|X] + 2E[(Y-Z)(Z-f(X))|X] + E[(Z-f(X))^2|X]. \end{aligned}$$

Now note that

$$E[(Y-Z)(Z-f(X))|X] = (Z-f(X))\underbrace{E[Y-Z|X]}_{0} = 0.$$

What now can you infer about $E[(Y-f(X))^2|X]$ and $E[(Y-Z)^2|X]$?

edited Sep 5 '16 at 15:46

answered Sep 5 '16 at 13:29
yurnero
7,827 1 9 26