Probability theory II Assignment 1

2021-21116 Taeyoung Chang

Exercises in Section 4.1 Conditional Expectation

4.1.1 **Bayes' formula.** Let $G \in \mathcal{G}$ and show that

$$P(G|A) = \int_G P(A|\mathcal{G})\,dP \Big/ \int_\Omega P(A|\mathcal{G})\,dP$$

When $\mathcal{G}$ is the σ-field generated by a partition, this reduces to the usual Bayes' formula

$$P(G_i|A) = P(A|G_i)P(G_i) \Big/ \sum_j P(A|G_j)P(G_j)$$

Remark. *We didn't define 'conditional prob. given a set' in the lecture, but in the textbook, the author said that "To continue the connection with undergraduate notions, $P(A|B) := P(A \cap B) \,/\, P(B)$"*

Proof. $P(G|A) = P(G \cap A)/P(A)$. Note that $P(A|\mathcal{G}) = E[I_A|\mathcal{G}]$

$$\int_G P(A|\mathcal{G})\,dP = \int_G E[I_A|\mathcal{G}]\,dP = \int_G I_A\,dP = P(A \cap G)$$
$$\int_\Omega P(A|\mathcal{G})\,dP = \int_\Omega E[I_A|\mathcal{G}]\,dP = \int_\Omega I_A\,dP = P(A)$$

The second equality in each of two equations above is due to the definition of conditional expectation and the fact that $G, \Omega \in \mathcal{G}$. Therefore the equaltiy $P(G|A) = \int_G P(A|\mathcal{G})\,dP / \int_\Omega P(A|\mathcal{G})\,dP$ holds true. To prove the equation for Bayes' formula, we shall use a lemma learned in the lecture.
(Lemma) $\Omega = \bigcup_{i=1}^\infty \Omega_i$ is a partition of Ω with $\Omega_i \in \mathcal{F}_0$ and $P(\Omega_i) > 0 \quad \forall\, i \in \mathbb{N}$
$\mathcal{F} = \sigma\{\Omega_1, \Omega_2, \cdots\} = \{\bigcup_{j \in \kappa} \Omega_j : \kappa \subset \mathbb{N}\}$ ($\mathcal{F}$ is a σ-field). Then we have

$$E[X|\mathcal{F}] = \sum_{i=1}^\infty a_i I_{\Omega_i} \quad \textit{with} \quad a_i = \frac{E[XI_{\Omega_i}]}{P(\Omega_i)}$$

Now, $\mathcal{G} = \sigma\{G_1, G_2, \cdots\}$ with $\Omega = \bigcup_{i=1}^\infty G_i$ is a partition.

$$\begin{aligned}
\int_{G_i} P(A|\mathcal{G})\,dP &= \int_{G_i} E[I_A|\mathcal{G}]\,dP \underset{\textit{lemma}}{=} \int_{G_i} \sum_{j=1}^\infty \frac{E[I_A I_{G_j}]}{P(G_j)} I_{G_j}\,dP \underset{MCT}{=} \sum_{j=1}^\infty \int_{G_i} \frac{E[I_{A \cap G_j}]}{P(G_j)} I_{G_j}\,dP \\
&\underset{\textit{partition}}{=} \int_{G_i} \frac{P(A \cap G_i)}{P(G_i)} \cdot 1\,dP = P(A|G_i)P(G_i) \\
\int_\Omega P(A|\mathcal{G})\,dP &= \int_\Omega E[I_A|\mathcal{G}]\,dP \underset{\textit{lemma}}{=} \int_\Omega \sum_{j=1}^\infty \frac{E[I_A I_{G_j}]}{P(G_j)} I_{G_j}\,dP \underset{MCT}{=} \sum_{j=1}^\infty \int_\Omega \frac{E[I_{A \cap G_j}]}{P(G_j)} I_{G_j}\,dP \\
&= \sum_{j=1}^\infty \int_{G_j} \frac{P(A \cap G_j)}{P(G_j)} \cdot 1\,dP = \sum_{j=1}^\infty P(A|G_j)P(G_j)
\end{aligned}$$

Therefore the equality $P(G_i|A) = P(A|G_i)P(G_i) \,/\, \sum_j P(A|G_j)P(G_j)$ holds true. $\square$

4.1.2 Prove **Chebyshev's inequality.** If $a > 0$, then

$$P(|X| \geq a|\mathcal{F}) \leq a^{-2}E(X^2|\mathcal{F})$$

Proof.

$$P(|X| \geq a \,|\, \mathcal{F}) = E[I(|X| \geq a)|\mathcal{F}] = E[I(X^2 \geq a^2)|\mathcal{F}] \leq E\Big[\frac{X^2}{a^2}I(X^2 \geq a^2)|\mathcal{F}\Big]$$
$$= \frac{1}{a^2}E[X^2 I(X^2 \geq a^2)|\mathcal{F}] \leq \frac{1}{a^2}E[X^2|\mathcal{F}]$$

□

4.1.3 Imitate the proof in the remark after Theorem 1.5.2 to prove the conditional Cauchy-Schwarz inequality.

$$E(XY|\mathcal{G})^2 \leq E(X^2|\mathcal{G})E(Y^2|\mathcal{G})$$

Proof. Take arbitrary $a \in \mathbb{R}$. Assume $E[X^2]$, $E[Y^2] < \infty$

$$0 \leq E[(aX+Y)^2|\mathcal{G}] = a^2E[X^2|\mathcal{G}] + 2aE[XY|\mathcal{G}] + E[Y^2|\mathcal{G}]$$
$$= E[X^2|\mathcal{G}]\Big(a + \frac{E[XY|\mathcal{G}]}{E[X^2|\mathcal{G}]}\Big)^2 + E[Y^2|\mathcal{G}] - \frac{E^2[XY|\mathcal{G}]}{E[X^2|\mathcal{G}]}$$

Since the inequality above holds for any $a \in \mathbb{R}$, $E[Y^2|\mathcal{G}] - \frac{E^2[XY|\mathcal{G}]}{E[X^2|\mathcal{G}]} \geq 0$ must hold, which implies that $E^2[XY|\mathcal{G}] \geq E[X^2|\mathcal{G}]E[Y^2|\mathcal{G}]$ □

4.1.4 Use regular conditional probability to get the conditional Hölder inequality from the unconditional one, i.e., show that if $p, q \in (1, \infty)$ with $1/p + 1/q = 1$, then

$$E(|XY||\mathcal{G}) \leq E(|X|^p|\mathcal{G})^{1/p}E(|Y|^q|\mathcal{G})^{1/q}$$

Proof. Assume $X \in \mathcal{L}^p$ and $Y \in \mathcal{L}^q$.
p and q are conjugate exponents. Then by Young's inequality,

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \quad \forall\, x, y > 0$$

Then for any $\varepsilon > 0$, we have

$$\frac{|X|}{(E[|X|^p|\mathcal{G}]+\varepsilon)^{1/p}} \cdot \frac{|Y|}{(E[|Y|^q|\mathcal{G}]+\varepsilon)^{1/q}} \leq \frac{1}{p}\frac{|X|^p}{E[|X|^p|\mathcal{G}]+\varepsilon} + \frac{1}{q}\frac{|Y|^q}{E[|Y|^q|\mathcal{G}]+\varepsilon} \quad \cdots (\bigstar)$$

Note that $E[|X|^p|\mathcal{G}]$ and $E[|Y|^q|\mathcal{G}]$ are $\mathcal{G}$-measurable. Also recall that if $Z > 0$ and Z is $\mathcal{G}$-measurable then $1/Z$ is also $\mathcal{G}$-measurable. Hence $1/(E[|X|^p|\mathcal{G}]+\varepsilon)$ and $1/(E[|Y|^q|\mathcal{G}]+\varepsilon)$ are all $\mathcal{G}$-measurable. We shall take $E[\,\cdot\,|\mathcal{G}]$ on both sides of inequality above. Before that, we should check that both sides

of the inequality ($\bigstar$) are integrable.
Note that XY is integrable by the assumption $X \in \mathcal{L}^p$, $Y \in \mathcal{L}^q$ and original Hölder inequality.

$$0 \le E[\,LHS\,] = E\Big[\frac{|X|}{(E[|X|^p|\mathcal{G}] + \varepsilon)^{1/p}} \cdot \frac{|Y|}{(E[|Y|^q|\mathcal{G}] + \varepsilon)^{1/q}}\Big] \le E\Big[\frac{|X|}{\varepsilon^{1/p}} \cdot \frac{|Y|}{\varepsilon^{1/q}}\Big] = \frac{1}{\varepsilon}E|XY| < \infty$$

$$0 \le E[\,RHS\,] = E\Big[\frac{1}{p}\frac{|X|^p}{E[|X|^p|\mathcal{G}] + \varepsilon} + \frac{1}{q}\frac{|Y|^q}{E[|Y|^q|\mathcal{G}] + \varepsilon}\Big] \le E\Big[\frac{1}{p}\frac{|X|^p}{\varepsilon} + \frac{1}{q}\frac{|Y|^q}{\varepsilon}\Big] = \frac{1}{\varepsilon}\Big(\frac{1}{p}E|X|^p + \frac{1}{q}E|Y|^q\Big) < \infty$$

Thus the both sides of inequality ($\bigstar$) are integrable. (Actually, we add ε for this.) Now take $E[\,\cdot\,|\mathcal{G}]$ on both sides. Then by smoothing property and linearity of conditional expectation, we have

$$\begin{aligned}\frac{1}{(E[|X|^p|\mathcal{G}] + \varepsilon)^{1/p}}\frac{1}{(E[|Y|^q|\mathcal{G}] + \varepsilon)^{1/q}}E[|XY|\,\big|\mathcal{G}] &\le \frac{1}{p}\frac{1}{E[|X|^p|\mathcal{G}] + \varepsilon}E[|X|^p|\mathcal{G}] + \frac{1}{q}\frac{1}{E[|Y|^q|\mathcal{G}] + \varepsilon}E[|Y|^q|\mathcal{G}] \\ &\le \frac{1}{p} + \frac{1}{q} = 1\end{aligned}$$

Hence we have

$$E[|XY|\,\big|\mathcal{G}] \le (E[|X|^p|\mathcal{G}] + \varepsilon)^{1/p}(E[|Y|^q|\mathcal{G}] + \varepsilon)^{1/q} \quad \forall\, \varepsilon > 0$$

Taking $\varepsilon \to 0$, we have $E[|XY|\,\big|\mathcal{G}] \le E[|X|^p|\mathcal{G}]^{1/p}E[|Y|^q|\mathcal{G}]^{1/q}$ as desired. □

Remark. *The problem was intended to use regular conditional probability in textbook, but I think this proof is more intuitive.*

4.1.5 Give an example on $\Omega = \{a, b, c\}$ in which

$$E(E(X|\mathcal{F}_1)|\mathcal{F}_2) \ne E(E(X|\mathcal{F}_2)|\mathcal{F}_1)$$

Proof. Set two σ-fields F_1, $F_2 \subset \mathcal{P}(\Omega)$ as below :

$$\mathcal{F}_1 = \sigma(\{a\}) = \big\{\phi, \{a\}, \{b, c\}, \{a, b, c\}\big\}$$
$$\mathcal{F}_2 = \sigma(\{c\}) = \big\{\phi, \{c\}, \{a, b\}, \{a, b, c\}\big\}$$

$\mathcal{F}_0 := \mathcal{P}(\Omega)$. $\mathcal{F}_1$ and $\mathcal{F}_2$ are sub σ-fields of $\mathcal{F}_0$. Define $X : \Omega \to \mathbb{R}$ by $X(a) = 0,\ X(b) = 1,\ X(c) = 0$

$$(X \ge k) = \begin{cases} \phi \in \mathcal{F}_0 & if \quad k > 1 \\ \{b\} \in \mathcal{F}_0 & if \quad 0 < k \le 1 \\ \{a, b, c\} \in \mathcal{F}_0 & if \quad k \le 0 \end{cases}$$

X is a F_0-measurable random variable. Notice that X is neither $\mathcal{F}_1$-measurable, nor $\mathcal{F}_2$-measurable. Also, we shall define probabiliy measure P on $(\Omega, \mathcal{F}_0)$ by $P(\{a\}) = P(\{b\}) = P(\{c\}) = 1/3$
Our strategy for the proof is that find out $E[X|\mathcal{F}_1]$ and $E[E[X|\mathcal{F}_1]|\mathcal{F}_2]$ in order, and then claim that it is not $\mathcal{F}_1$-measurable so that it cannot be equal to $E[E[X|\mathcal{F}_2]|\mathcal{F}_1]$.

i. Find out $E[X|\mathcal{F}_1]$
   In $\mathcal{F}_1$, unlike in $\mathcal{F}_0$, $\{b\}$ and $\{c\}$ cannot be separated. Since $P(\{b\}) = P(\{c\}) = 1/3$ and $(X(b) + X(c))/2 = 1/2$, we can guess that Y defined by $Y(a) = 0,\ Y(b) = 1/2,\ Y(c) = 1/2$ might be $E[X|\mathcal{F}_1]$.

$$(Y \ge k) = \begin{cases} \phi \in \mathcal{F}_1 & if \quad k > 1/2 \\ \{b, c\} \in \mathcal{F}_1 & if \quad 0 < k \le 1/2 \\ \{a, b, c\} \in \mathcal{F}_1 & if \quad k \le 0 \end{cases}$$

Hence Y is $\mathcal{F}_1$-measurable.

$$\int_{\phi} X\,dP = 0 \qquad \int_{\{b,c\}} X\,dP = 0\cdot P(\{b\}) + 1\cdot P(\{c\}) = 1/3$$
$$\int_{\{a\}} X\,dP = 0\cdot P(\{a\}) = 0 \qquad \int_{\{a,b,c\}} X\,dP = 0\cdot P(\{a\}) + 1\cdot P(\{b\}) + 0\cdot P(\{c\}) = 1/3$$
$$\int_{\phi} Y\,dP = 0 \qquad \int_{\{b,c\}} Y\,dP = 1/2\cdot P(\{b\}) + 1/2\cdot P(\{c\}) = 1/3$$
$$\int_{\{a\}} Y\,dP = 0\cdot P(\{a\}) = 0 \qquad \int_{\{a,b,c\}} Y\,dP = 0\cdot P(\{a\}) + 1/2\cdot P(\{b\}) + 1/2\cdot P(\{c\}) = 1/3$$

Thus $\int_A X\,dP = \int_A Y\,dP \quad \forall\, A \in \mathcal{F}_1. \quad Y = E[X|\mathcal{F}_1]\ a.s.$

ii. Find out $E[E[X|\mathcal{F}_1]|\mathcal{F}_2]$
In $\mathcal{F}_2$, $\{a\}$ and $\{b\}$ cannot be separated. Since $P(\{a\}) = P(\{b\})$ and $(Y(a)+Y(b))/2 = 1/4$, we can guess that Z defined by $Z(a) = 1/4,\ Z(b) = 1/4,\ Z(c) = 1/2$ might be $E[Y|\mathcal{F}_2] = E[E[X|\mathcal{F}_1]|\mathcal{F}_2]$.

$$(Z \geq k) = \begin{cases} \phi \in \mathcal{F}_2 & if \quad k > 1/2 \\ \{c\} \in \mathcal{F}_2 & if \quad 1/4 < k \leq 1/2 \\ \{a,b,c\} \in \mathcal{F}_2 & if \quad k \leq 1/4 \end{cases}$$

Hence Z is $\mathcal{F}_2$-measurable.

$$\int_{\phi} Y\,dP = 0 \qquad \int_{\{a,b\}} Y\,dP = 0\cdot P(\{a\}) + 1/2\cdot P(\{b\}) = 1/6$$
$$\int_{\{c\}} Y\,dP = 1/2\cdot P(\{c\}) = 1/6 \qquad \int_{\{a,b,c\}} Y\,dP = 0\cdot P(\{a\}) + 1/2\cdot P(\{b\}) + 1/2\cdot P(\{c\}) = 1/3$$
$$\int_{\phi} Z\,dP = 0 \qquad \int_{\{a,b\}} Z\,dP = 1/4\cdot P(\{a\}) + 1/4\cdot P(\{b\}) = 1/6$$
$$\int_{\{c\}} Z\,dP = 1/2\cdot P(\{c\}) = 1/6 \qquad \int_{\{a,b,c\}} Z\,dP = 1/4\cdot P(\{a\}) + 1/4\cdot P(\{b\}) + 1/2\cdot P(\{c\}) = 1/3$$

Thus $\int_A Y\,dP = \int_A Z\,dP \quad \forall\, A \in \mathcal{F}_2. \quad Z = E[Y|\mathcal{F}_2] = E[E[X|\mathcal{F}_1]|\mathcal{F}_2]$

iii. Conclusion
Observe that Z is not $\mathcal{F}_1$-measurable. This is because $(Z \geq k) = \{c\} \notin \mathcal{F}_1$ whenever $1/4 < k \leq 1/2$. Suppose $E[E[X|\mathcal{F}_1]|\mathcal{F}_2] = E[E[X|\mathcal{F}_2]|\mathcal{F}_1]\ a.s.$. Then since $P(\{a\}) = P(\{b\}) = P(\{c\}) = 1/3 > 0$, $\ E[E[X|\mathcal{F}_2]|\mathcal{F}_1] = Z$ (exactly same). This means that $E[E[X|\mathcal{F}_2]|\mathcal{F}_1]$ is not $\mathcal{F}_1$-measurable, which is a contradiction.
Therefore, $E[E[X|\mathcal{F}_1]|\mathcal{F}_2] \neq E[E[X|\mathcal{F}_2]|\mathcal{F}_1]$ in this case.

$\square$

4.1.6 Show that if $\mathcal{G} \subset \mathcal{F}$ and $EX^2 < \infty$, then

$$E(\{X - E(X|\mathcal{F})\}^2) + E(\{E(X|\mathcal{F}) - E(X|\mathcal{G})\}^2) = E(\{X - E(X|\mathcal{G})\}^2)$$

Dropping the second term on the left, we get an inequality that says geometrically, the larger the subspace the closer the projection is, or statistically, more information means a smaller mean square error.

Remark. $E[\{X - E[X|\mathcal{F}]\}^2] \leq E[\{X - E[X|\mathcal{G}]\}^2]$ *whenever* $\mathcal{G} \subset \mathcal{F}$, *provided* $X \in \mathcal{L}^2$
The more information given, the closer to the target random variable in $\mathcal{L}^2$ *distance.*

Proof.

$$\begin{aligned} E[\{X - E[X|\mathcal{G}]\}^2] &= E[\{X - E[X|\mathcal{F}] + E[X|\mathcal{F}] - E[X|\mathcal{G}]\}^2] \\ &= E[\{X - E[X|\mathcal{F}]\}^2] + E[\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}^2] + 2E[\{X - E[X|\mathcal{F}]\}\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}] \end{aligned}$$

Second equality holds because $X \in \mathcal{L}^2 \Rightarrow E[X|\mathcal{F}], E[X|\mathcal{G}] \in \mathcal{L}^2$ and $\mathcal{L}^2$ is a vector space. Now, it suffices to show that the cross product $E[\{X - E[X|\mathcal{F}]\}\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}]$ is zero. We shall take advantage of a simple fact that $\mathcal{G} \subset \mathcal{F} \Rightarrow$ Every $\mathcal{G}$-measurable r.v. is $\mathcal{F}$-measurable. $\cdots (*)$

$$\begin{aligned} CP &= E[\{X - E[X|\mathcal{F}]\}\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}] \\ &= E[E[\{X - E[X|\mathcal{F}]\}\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}|\mathcal{F}]] \quad \because E[Z] = E[E[Z|\mathcal{F}]] \\ &= E[\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}E[\{X - E[X|\mathcal{F}]\}|\mathcal{F}]] \quad \because E[X|\mathcal{F}] - E[X|\mathcal{G}] \in \mathcal{F} \textit{ by } (*) \\ &= E[0] \quad \because E[\{X - E[X|\mathcal{F}]\}|\mathcal{F}] = E[X|\mathcal{F}] - E[X|\mathcal{F}]E[1|\mathcal{F}] = E[X|\mathcal{F}] - E[X|\mathcal{F}] = 0 \\ &= 0 \end{aligned}$$

Therefore $E[\{X - E[X|\mathcal{G}]\}^2] = E[\{X - E[X|\mathcal{F}]\}^2] + E[\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}^2]$ $\square$

4.1.7 An important special case of the previous result occurs when $\mathcal{G} = \{\emptyset, \Omega\}$. Let $\operatorname{var}(X|\mathcal{F}) = E(X^2|\mathcal{F}) - E(X|\mathcal{F})^2$. Show that

$$\operatorname{var}(X) = E(\operatorname{var}(X|\mathcal{F})) + \operatorname{var}(E(X|\mathcal{F}))$$

Remark. $Var(X|\mathcal{F}) := E[X^2|\mathcal{F}] - E^2[X|\mathcal{F}] = E[\{X - E[X|\mathcal{F}]\}^2|\mathcal{F}]$

Proof. Note that $E[X|\mathcal{G}] = E[X]$ when $\mathcal{G} = \{\phi, \Omega\}$. We shall plug in $\mathcal{G} = \{\phi, \Omega\}$ on the equation derived at problem 4.1.6

i. $E[\{X - E[X|\mathcal{G}]\}^2] = E[\{X - E(X)\}^2] = Var(X)$

ii. $E[\{X - E[X|\mathcal{F}]\}^2] = E[E[\{X - E[X|\mathcal{F}]\}^2|\mathcal{F}]] = E[Var(X|\mathcal{F})]$

iii. $E[\{E[X|\mathcal{F}] - E[X|\mathcal{G}]\}^2] = E[\{E[X|\mathcal{F}] - E[X]\}^2] = E[\{E[X|\mathcal{F}] - E[E[X|\mathcal{F}]]\}^2] = Var(E[X|\mathcal{F}])$

Problem 4.1.6 tells us that i = ii+ iii. Thus $Var(X) = E[Var(X|\mathcal{F})] + Var(E[X|\mathcal{F}])$ $\square$

4.1.9 Show that if X and Y are random variables with $E(Y|\mathcal{G}) = X$ and $EY^2 = EX^2 < \infty$, then $X = Y$ a.s.

Proof. We again use the equation derived at problem 4.1.6 with plugging in trivial σ-field (which is a trick also used for problem 4.1.7)

$$E[\{Y - E[Y|\mathcal{F}]\}^2] = E[\{Y - E[Y|\mathcal{G}]\}^2] + E[\{E[Y|\mathcal{G}] - E[Y|\mathcal{F}]\}^2] \quad \textit{whenever } \mathcal{F} \subset \mathcal{G},\ E[Y^2] < \infty$$

Now plug in $\mathcal{F} = \{\phi, \Omega\}$. Then we have

i. $E[\{Y - E[Y|\mathcal{F}]\}^2] = E[\{Y - E[Y]\}^2] = Var(Y)$

ii. $E[\{Y - E[Y|\mathcal{G}]\}^2] = E[\{Y - X\}^2]$

iii. $E[\{E[Y|\mathcal{G}] - E[Y|\mathcal{F}]\}^2] = E[\{X - E[Y]\}^2]$

Thus we have $Var(Y) = E[\{Y - X\}^2] + E[\{X - E[Y]\}^2] \ \cdots\ (\star)$
Notice that $E[X] = E[E[Y|\mathcal{G}]] = E[Y]$. $(\star)$ turns to $Var(Y) = E[\{Y - X\}^2] + Var(X)$.
Since $E(X) = E(Y)$ and $E(X^2) = E(Y^2) < \infty$ by assumption, we have $Var(X) = Var(Y) < \infty$
Therefore, we get $E[\{Y - X\}^2] = 0$. It implies that $(Y - X)^2 = 0\ \ a.s. \quad \because (Y - X)^2 \geq 0$
Thus $|Y - X| = 0\ \ a.s.$ and $X = Y\ a.s.$ $\square$