Statistics 522, Problem Set 2, Solutions

Wellner; 1/23/2013

1. Show that $Z = E(Y|\mathcal{D})$ minimizes $E(Y-Z)^2$ among all $\mathcal{D}-$measurable random variables $Z \in \mathcal{H}_{\mathcal{D}}$.

   Solution: Let $Z \in \mathcal{H}_{\mathcal{D}}$, and note that

$$
\begin{aligned}
E(Y-Z)^2 &= E(Y - E(Y|\mathcal{D}) + E(Y|\mathcal{D}) - Z)^2 \\
&= E(Y - E(Y|\mathcal{D}))^2 + E\{(Y - E(Y|\mathcal{D}))(E(Y|\mathcal{D}) - Z)\} \\
&\quad + E\{(E(Y|\mathcal{D}) - Z)^2\} \\
&= E(Y - E(Y|\mathcal{D}))^2 + 0 + E\{(E(Y|\mathcal{D}) - Z)^2\} \\
&\geq E(Y - E(Y|\mathcal{D}))^2
\end{aligned}
$$

   where the 0 for the cross-term in the display holds by computing conditionally on $\mathcal{D}$:

$$
\begin{aligned}
E\{(Y - E(Y|\mathcal{D}))(E(Y|\mathcal{D}) - Z)\} &= EE\{(Y - E(Y|\mathcal{D}))(E(Y|\mathcal{D}) - Z)|\mathcal{D}\} \\
&= E\left\{(E(Y|\mathcal{D}) - Z)E\{Y - E(Y|\mathcal{D})|\mathcal{D}\}\right\} \\
&= E\left\{(E(Y|\mathcal{D}) - Z) \cdot 0\right\} \\
&= 0
\end{aligned}
$$

   where the $\mathcal{D}-$ measurability of $E(Y|\mathcal{D}) - Z$ was used in the second equality.

2. Suppose that Y is a random variable defined on $(\Omega, \mathcal{A}, P)$ and that $EY^2 < \infty$. Moreover, suppose $\mathcal{D} \subset \mathcal{A}$ is a sub-$\sigma-$field of $\mathcal{A}$.
   (a) Show that $Var(Y) = Var(E(Y|\mathcal{D})) + E\{Var(Y|\mathcal{D})\}$.
   (b) Can you relate this to the first problem above?

   Solution: (a) Using the same argument as in problem 1 to get the third equality,

$$
\begin{aligned}
Var(Y) &= E(Y - EY)^2 = E\left(Y - E(Y|\mathcal{D}) + E(Y|\mathcal{D}) - E(Y)\right)^2 \\
&= E\left(Y - E(Y|\mathcal{D})\right)^2 + E\{(Y - E(Y|\mathcal{D}))(E(Y|\mathcal{D}) - E(Y))\} \\
&\quad + E\left(E(Y|\mathcal{D}) - E(Y)\right)^2 \\
&= E\left(Y - E(Y|\mathcal{D})\right)^2 + 0 + E\left(E(Y|\mathcal{D}) - E(Y)\right)^2 \\
&= E\mathrm{Var}(Y|\mathcal{D}) + Var(E(Y|\mathcal{D})).
\end{aligned}
$$

(b) By problem 1, $E(Y-Z)^2$ is minimized over $\mathcal{D}-$measurable variables Z by $E(Y|\mathcal{D})$, and then part (a) above gives a Pythagorean relationship for the (orthogonal) decomposition $Y - EY = (Y - E(Y|\mathcal{D})) + (E(Y|\mathcal{D}) - EY)$ in $L_2(P)$.

3. PfS, Exercise 7.5.3, PfS Course Notes, page 145:
Theorem 5.4. Suppose that $X : (\Omega, \mathcal{A}, P) \to (M_1, \mathcal{G}_1)$ and $Y : (\Omega, \mathcal{A}, P) \to (M_2, \mathcal{G}_2)$ where $(M_1, \mathcal{G}_1)$ and $(M_2, \mathcal{G}_2)$ are Borel spaces. Then $(X, Y) : (\Omega, \mathcal{A}, P) \to (M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2)$. Furthermore

$$\text{a regular conditional probability } \; P(A|X = x) \text{ exists}$$

for sets $A \in \tilde{\mathcal{A}} \equiv Y^{-1}(\mathcal{G}_2) \subset \mathcal{A}$ and for $x \in M_1$. Let $E|h(X,Y)| < \infty$.
(a) Then

$$E\{h(X,Y)|X = x\} = \int_{M_2} h(x,y) dP(y|X = x) \;\; \text{a.s.}$$

(b) If X and Y are independent, then

$$E\{h(X,Y)|X = x\} = E\{h(x,Y)\} \;\; \text{a.s.}$$

Solution: First, $(M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2)$ is a Borel space. To see this, define $\phi : (M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2) \to ((0,1), \mathcal{B} \cap (0,1))$ by $\phi(m_1, m_2) = h \circ g \circ f(m_1, m_2)$ where $f(m_1, m_2) = (\phi_1(m_1), \phi_2(m_2))$ is a one-to-one bimeasurable map from $M_1 \times M_2$ to $\mathbb{R}^2$, $g(x_1, x_2) = (\Phi(x_1), \Phi(x_2))$ with Φ the standard normal distribution function is a one-to-one bimeasurable map from $\mathbb{R}^2$ to $(0,1)^2$, and $h(y_1, y_2) = .y_{11}y_{21}y_{12}y_{22}y_{13}y_{23}\cdots$ with $y_1 = .y_{11}y_{12}y_{13}\cdots$ and $y_2 = .y_{21}y_{22}y_{23}\cdots$ is a one-to-one bimeasurable map from $(0,1)^2$ to $(0,1)$. Thus ϕ gives a one-to-one bimeasurable map, and the first claim follows.

Now note that $(X,Y) : (\Omega, \mathcal{A}, P) \mapsto (M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2)$ is measurable and induces a probability measure $P_{(X,Y)}$ on $(M_1 \times M_2, \mathcal{G}_1 \times \mathcal{G}_2)$. Thus with $\mathcal{D} \equiv X^{-1}(\mathcal{G}_1)$ and $Z \equiv h(X,Y)$ for any measurable map to some Borel space $(M, \mathcal{G})$ we know that $P_Z^\omega(\cdot|\mathcal{D})$ exists as a regular probability measure on $\mathcal{G}$. In particular, with $Z \equiv h(X,Y) = Y$ and $(M, \mathcal{G}) = (M_2, \mathcal{G}_2)$,

$$P_Y^\omega(G_2|\mathcal{D}) = P_Y(G_2|X)(\omega)$$

exists as a regular conditional probability for $G_2 \in \mathcal{G}_2$. Since this is measurable with respect to X, we can then define it in terms of the measurable function

$$P_Y^\omega(G_2|\mathcal{D}) = P_Y(G_2|X)(\omega) \equiv \xi(X(\omega))$$

where $\xi(m_1) \equiv P_Y^\omega(G_2|x = m_1)$.

Now we want to show that

$$E\{h(X,Y)|X = x\} = \int_{M_2} h(x,y)dP_Y(y|X = x).$$

This holds if and only if

$$\begin{aligned} &\int_{M_1} 1_D(x)\left(\int_{M_2} h(x,y)dP_Y(y|X = x)\right)dP_X(x) \\ &= \int_{M_1\times M_2} 1_D(x)h(x,y)dP_{(X,Y)}(x,y) \end{aligned} \tag{1}$$

for all $D \in \mathcal{G}_1$.

To prove this, first consider $h(x,y) = 1_{G_1}(x)1_{G_2}(y)$ for $G_j \in \mathcal{G}_j$, $j = 1, 2$. Then we have

$$\begin{aligned} &\int_{M_1} 1_D(x)\left(\int_{M_2} 1_{G_1}(x)1_{G_2}(y)dP_Y(y|X = x)\right)dP_X(x) \\ &= \int_{M_1} 1_D(x)1_{G_1}(x)\int_{M_2} 1_{G_2}(y)dP_Y(y|X = x)dP_X(x) \\ &= \int_{M_1} 1_{D\cap G_1}(x)P_Y(G_2|X = x)dP_X(x) \\ &= P_{X,Y}((X,Y) \in (D \cap G_1) \times G_2) \quad \text{by definition of} \quad P_Y(G_2|X = x) \\ &= \int_{M_1\times M_2} 1_D(x)1_{G_1}(x)1_{G_2}(y)dP_{(X,Y)}(x,y), \end{aligned}$$

so the identity holds in this case. Now for $D \in \mathcal{G}_1$ fixed, define two measures ν_1 and ν_2 on $\mathcal{G} = \mathcal{G}_1 \times \mathcal{G}_2$ by:

$$\begin{aligned} \nu_1(G) &= \int_{M_1} 1_D(x)\left(\int_{M_2} 1_G(x,y)dP_Y(y|X = x)\right)dP_X(x), \\ \nu_2(G) &= \int\int_{M_1\times M_2} 1_D(x)1_G(x,y)dP_{X,Y}(x,y). \end{aligned}$$

Now ν_1 and ν_2 agree on the $\overline{\pi}-$system $\{G_1 \times G_2 : \ G_j \in \mathcal{G}_j, \ j = 1, 2\}$. Hence by the $\pi - \lambda$ theorem they agree on $\mathcal{G} \equiv \mathcal{G}_1 \times \mathcal{G}_2$.

Next consider simple functions $h(x,y) = \sum_{i=1}^{k} c_i 1_{G_i}(x,y)$ for $G_i \in \mathcal{G}$ and c_i real. Then (1) holds by the result of the last paragraph together with linearity of the integrals. Now consider $h \geq 0$ and measurable. Then there exist simple functions $h_k \nearrow h$, and we see that the key identity holds by virtue of the result for simple functions and the (conditional) monotone convergence theorem on one side and the MCT on the other side. Finally, consider $h = h^+ - h^-$. Then (1) holds by virtue of its validity for $h \geq 0$ and linearity.

To see that the claim in (b) holds, we need to show that

$$\int_{M_1} 1_D(x) E\{h(x,Y)\} dP_X(x) = \int\int_{M_1 \times M_2} 1_D(x) h(x,y) dP_{X,Y} \qquad (2)$$

for all $D \in \mathcal{G}_1$. But for X, Y independent $P_{X,Y}((X,Y) \in D_1 \times D_2) = P_X(X \in D_1) P_Y(Y \in D_2)$ and hence the right side in (2) becomes

$$\begin{aligned} &\int\int_{M_1 \times M_2} 1_D(x) h(x,y) dP_X(x) dP_Y(y) \\ &= \int_{M_1} 1_D(x) \left(\int_{M_2} h(x,y) dP_Y(y) \right) dP_X(x) \\ &= \int_{M_1} 1_D(x) E\{h(x,Y)\} dP_X(x) \end{aligned}$$

by Fubini's theorem.

4. Suppose that $X, Y \in L_1(\Omega, \mathcal{F}, P)$ and that $E(Y|X) = X$ a.s. and $E(X|Y) = Y$ a.s. Prove that $P(X = Y) = 1$. (See e.g. Exercise 9.2, Williams, *Probability with Martingales*, page 231.)

   Solution: Suppose first that $X, Y \in L_2(P)$. Then, by Pythagoras,

   $$EX^2 = E(E(X|Y)^2) + E((X - E(X|Y))^2),$$

   and since $E(X|Y) = Y$ a.s. this yields

   $$E(X^2) = E(Y^2) + E((X - Y)^2). \qquad (3)$$

Reversing the roles of X and Y, we also obtain, upon using $E(Y|X) = X$ a.s.,

$$E(Y^2) = E(X^2) + E((Y - X)^2)\,. \tag{4}$$

Adding (3) and (4) gives

$$E(X^2) + E(Y^2) = E(X^2) + E(Y^2) + 2E((X - Y)^2)\,,$$

and this implies that $E(X - Y)^2 = 0$, which in turn implies $P(X = Y) = 1$.

Now one way to proceed is to reduce the general case of $X, Y \in L_1(P)$ to the $L_2(P)$ case treated above. Instead I will prove it using the hint.

Note that

$$\begin{aligned}
&E(X - Y)1_{[X>c,Y\le c]} + E(X - Y)1_{[X\le c,Y\le c]} \\
&\quad = E(X - Y)1_{[Y\le c]} = E(X1_{[Y\le c]}) - E(Y1_{[Y\le c]}) \\
&\quad = E(E(X1_{[Y\le c]}|Y)) - E(Y1_{[Y\le c]}) \\
&\quad = E(1_{[Y\le c]}E(X|Y)) - E(Y1_{[Y\le c]}) \\
&\quad = E(1_{[Y\le c]}Y) - E(Y1_{[Y\le c]}) = 0
\end{aligned} \tag{5}$$

using $E(X|Y) = Y$ a.s. in the last line. Similarly, reversing the roles of X and Y,

$$\begin{aligned}
&E(Y - X)1_{[Y>c,X\le c]} + E(Y - X)1_{[Y\le c,X\le c]} \\
&\quad = E(Y - X)1_{[X\le c]} = 0\,.
\end{aligned} \tag{6}$$

Adding (5) and (6) yields

$$\begin{aligned}
0 &= E(X - Y)1_{[X>c,Y\le c]} + E(X - Y)1_{[X\le c,Y\le c]} \\
&\qquad - E(X - Y)1_{[Y>c,X\le c]} - E(X - Y)1_{[Y\le c,X\le c]} \\
&= E(X - Y)1_{[X>c,Y\le c]} - E(X - Y)1_{[Y>c,X\le c]}\,.
\end{aligned}$$

Since $[X - Y > 0] = [X > Y] = \cup_{q\in\mathbf{Q}}[X > q \ge Y]$ and similarly for $[X - Y < 0]$, this yields, by summing over rationals q,

$$0 = E(X - Y)1_{[X-Y>0]} - E(X - Y)1_{[X-Y<0]} = E|X - Y|\,.$$

But this implies $P(|X - Y| = 0) = 1$, or $P(X = Y) = 1$.

5. (Symmetry and conditional expectation). Let $X_1, X_2, \ldots$ be i.i.d. random variables with the same distribution as X where $E|X| < \infty$. Let $S_n \equiv X_1 + \cdots + X_n$, and define

$$\mathcal{G}_n \equiv \sigma\left[S_n, S_{n+1}, \ldots\right] = \sigma\left[S_n, X_{n+1}, X_{n+2}, \ldots\right].$$

Show that $E(X_1|\mathcal{G}_n) = E(X_1|S_n) = n^{-1}S_n$ almost surely. [Hint: Note that $\sigma\left[X_{n+1}, X_{n+2}, \ldots\right]$ is independent of $\sigma\left[X_1, S_n\right]$, and use symmetry to show that $E(1_{[S_n \in B]}X_1) = E(1_{[S_n \in B]}X_2) = \cdots = E(1_{[S_n \in B]}X_n)$.]

Solution: Note that $\mathcal{T}_n = \sigma\{S_n, S_{n+1}, \ldots\} = \sigma\{S_n, X_{n+1}, X_{n+2}, \ldots\}$ and that $\sigma\{X_{n+1}, X_{n+2}, \ldots\}$ is independent of $\sigma\{X_1, S_n\}$. Thus by Theorem 7.4.1 part (23) it follows that

$$E(X_1|\mathcal{T}_n) = E(X_1|S_n).$$

But with $X_1 \sim F$, for any Borel set B we have

$$\begin{aligned} E\{1_B(S_n)X_1\} &= \int \cdots \int_{s_n \in B} x_1 dF(x_1) dF(x_2) \cdots dF(x_n) \\ &= E\{1_B(S_n)X_2\} = \cdots = E\{1_B(S_n)X_n\}. \end{aligned}$$

Hence

$$\begin{aligned} E(X_1|S_n) &= E(X_2|S_n) = \cdots = E(X_n|S_n) \qquad a.s. \\ &= n^{-1}E(X_1 + \cdots + X_n|S_n) = n^{-1}E(S_n|S_n) \\ &= n^{-1}S_n \qquad a.s.. \end{aligned}$$