Statistics 522, Problem Set 3 Solutions

Wellner; 2/05/2020

1. Suppose that X and Y are independent random variables with common continuous d.f. F. Let $M \equiv X \vee Y \equiv \max\{X, Y\}$. Show that for each fixed $x \in \mathbb{R}$

$$P(X \le x|M) = 1_{[M \le x]} + \frac{1}{2}\frac{F(x)}{F(M)} 1_{[M > x]}. \tag{1}$$

Hint: Since $\mathcal{M} = \{M \le m : m \in \mathbb{R}]\}$ is a $\overline{\pi}-$ family of sets in $\sigma[M]$, it suffices to show that

$$E\left\{1_{[M \le m]} P(X \le x|M)\right\} = P([M \le m] \cap [X \le x]) = P(M \le m, X \le x)$$

for all $m \in \mathbb{R}$.

Solution: To show that (1) holds we need to show that

$$E\{1_D P(X \le x|M)\} = P(D \cap [X \le x])$$

for all $D \in \sigma[M]$. Since the collection of sets $\{[M \le m] : \ m \in \mathbb{R}\}$ forms a $\overline{\pi}-$system, it is enough to show that for each $m \in \mathbb{R}$,

$$E\left\{1_{[M \le m]} P(X \le x|M)\right\} = P([M \le m] \cap [X \le x]).$$

That is, we want to show that

$$E\left\{1_{[M \le m]}\left(1_{[M \le x]} + \frac{1}{2}\frac{F(x)}{F(M)} 1_{[M > x]}\right)\right\} = P(M \le m, X \le x). \tag{2}$$

Now the left side in the last display can be written as

$$\begin{aligned} & P(M \le m \wedge x) + \frac{1}{2} E\left\{1_{[x < M \le m]} \frac{F(x)}{F(M)}\right\} 1_{[x < m]} \\ = \quad & P(M \le m \wedge x) + \frac{F(x)}{2} E\left\{1_{[x < M \le m]} \frac{1}{F(M)}\right\} 1_{[x < m]}. \end{aligned}$$

But, for $x < m$,

$$
\begin{aligned}
E\left\{1_{[x<M\le m]}\frac{1}{F(M)}\right\} &= \iint_{u\le v, x<v\le m}\frac{1}{F(v)}dF(u)dF(v) \\
&\quad + \iint_{v<u, x<u\le m}\frac{1}{F(u)}dF(u)dF(v) \\
&= \int_{x<v\le m} dF(v) + \int_{x<u\le m} dF(u) \\
&= 2(F(m)-F(x))
\end{aligned}
$$

where we used continuity of F. Thus we want to show that

$$P(M \le m\wedge x) + F(x)(F(m)-F(x))1_{[x<m]} = P(M\le m, X\le x).$$

When $x \ge m$ this becomes

$$P(M\le m) + 0 = P(M \le m)$$

so the identity (2) holds. When $x < m$ it becomes

$$
\begin{aligned}
P(M\le x) + F(x)F(m) - F^2(x) &= F^2(x) + F(x)F(m) - F^2(x) \\
&= F(x)F(m) = P(Y\le m, X\le x) \\
&= P(M\le m, X\le x),
\end{aligned}
$$

and hence (2) holds in this case as well.

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 $\mathcal{A}$, and let $Var(Y|\mathcal{D}) \equiv E\{[Y-E(Y|\mathcal{D})]^2|\mathcal{D}\} = E(Y^2|\mathcal{D}) - [E(Y|\mathcal{D})]^2$.
(a) Show that $Var(Y) = Var\{E(Y|\mathcal{D})\} + E\{Var(Y|\mathcal{D})\}$.
(b) Show that $Z\equiv E(Y|\mathcal{D})$ minimizes $E(Y-Z)^2$ over all $\mathcal{D}-$measurable random variables Z with $E(Z^2)<\infty$.

Solution: (a) Note that

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

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

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

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

(b) Let $Z \in \mathcal{H}_{\mathcal{D}}$. Much as in (a), 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 $\mathcal{D}$:

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

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

3. Exercise 7.4.1, page 131, PfS: show that if $\Omega=\sum_i D_i$ for a finite or countable collection of sets D_i, and if $\mathcal{D}\equiv\sigma[D_1,D_2,\ldots]$, then we can take

$$P(A|\mathcal{D})=\sum_i \frac{P(AD_i)}{P(D_i)}1_{D_i} \tag{3}$$

where $P(AD_i)/P(D_i)\equiv P(A)$ if $P(D_i)=0$. Also show that for general $Y\in\mathcal{L}_1$ we can take

$$E(Y|\mathcal{D})=\sum_i\left\{\frac{1}{P(D_i)}\int_{D_i}YdP\right\}1_{D_i}. \tag{4}$$

Solution: We need to show that the quantity on the right side of (3) satisfies

$$E\{1_D P(A|\mathcal{D})\} = E\{1_{D\cap A}\} \qquad \text{for all} \;\; D \in \mathcal{D}.$$

For $P(A|\mathcal{D})$ as defined in (3) and $B \in \mathcal{D}$ let

$$\nu_1(B) \equiv E\{1_B P(A|\mathcal{D})\},$$
$$\nu_2(B) \equiv E\{1_B 1_A\}.$$

With this notation we need to show that $\nu_1(B) = \nu_2(B)$ for all $B \in \mathcal{D}$. But since $\mathcal{D}$ is generated by the sets D_i in the $\bar{\pi}-$system $\{D_i\}$, it suffices, by Dynkin's $\pi - \lambda$ theorem, to show that $\nu_1(D_j) = \nu_2(D_j)$ for all j. But

$$\begin{aligned}
\nu_1(D_j) &= E\{1_{D_j} \sum_i \frac{P(AD_i)}{P(D_i)} 1_{D_i}\} \\
&= \sum_i \frac{P(AD_i)}{P(D_i)} P(D_j D_i) \\
&= \frac{P(AD_j)}{P(D_j)} P(D_j) \qquad \text{since} \;\; D_j D_i = \emptyset \;\; \text{for } i \neq j \\
&= P(AD_j) = E\{1_{D_j} 1_A\} = \nu_2(D_j).
\end{aligned}$$

Thus the right side of (3) is a version of $P(A|\mathcal{D})$.

To see that the right side of (4) is a version of $E(Y|\mathcal{D})$ in this case, we need to show that

$$E\{1_D E(Y|\mathcal{D})\} = E\{1_D Y\} \qquad \text{for all} \;\; D \in \mathcal{D}. \tag{5}$$

As above it suffices to check this for $D_j \in \{D_i\}$. But then

$$\begin{aligned}
& E\left\{1_{D_j} \sum_i \left\{\frac{1}{P(D_i)} \int_{D_i} Y dP\right\} 1_{D_i}\right\} \\
&= \sum_i \left\{\frac{1}{P(D_i)} \int_{D_i} Y dP\right\} P(D_j \cap D_i) \\
&= \left\{\frac{1}{P(D_j)} \int_{D_j} Y dP\right\} P(D_j) \qquad \text{since} \;\; D_j \cap D_i = \emptyset \;\; \text{for} \;\; i \neq j \\
&= E\{1_{D_j} Y\}.
\end{aligned}$$

Thus (5) holds and the proof is complete.

4. Exercise 7.4.5, page 139, PfS (2012). If X and Y are independent random variables with mean $\mu_Y = 0$, then for each $r \geq 1$ we have $E|X|^r \leq E|X+Y|^r$. More generally $E|X+\mu_Y|^r \leq E|X+Y|^r$.

   Solution: Note that $\mu_Y = E(Y) = E(Y|X)$ by independence of X and Y. Then since $X + E(Y|X) = E(X+Y|X)$ and the conditional version of Jensen's inequality for the convex function $g(z) = |z|^r$,

   $$|X+\mu_Y|^r = |X+E(Y|X)|^r = |E(X+Y|X)|^r \leq E\left\{|X+Y|^r|X\right\} \qquad \text{a.s.}.$$

   But then by monotonicity of expectation

   $$E|X+\mu_Y|^r \leq E\left[E\left\{|X+Y|^r|X\right\}\right] = E\{|X+Y|^r\}.$$

5. Exercise 7.4.4, page 139, PfS (2012). (In proving the statement (26), page 136, it is to be understood that $E(XY)$ exists; alternatively, show that the statement holds for all *bounded* $\mathcal{D}-$measurable random variables X.)

   Solution: (24): C_r: For $r \geq 1$, $|x|^r$ is a convex function of x, so $|(x+y)/2|^r \leq (1/2)(|x|^r + |y|^r)$. Thus $|X+Y|^r \leq 2^{r-1}\{|X|^r + |Y|^r\}$. Taking condition expectations across this inequality and using (16) yields $E(|X+Y|^r|\mathcal{D}) \leq 2^{r-1}\{E(|X|^r|\mathcal{D}) + E(|Y|^r|\mathcal{D})\}$. For $0 < r \leq 1$, $|X+Y|^r \leq |X|^r + |Y|^r$, so taking conditional expectations across this inequality yields $E(|X+Y|^r|\mathcal{D}) \leq E(|X|^r|\mathcal{D}) + E(|Y|^r|\mathcal{D})$.
   Hölder's inequality: for arbitrary $a, b \in R$ and r, s satisfying $(1/r) + (1/s) = 1$, we have

   $$|ab| \leq \frac{|a|^r}{r} + \frac{|b|^s}{s}$$

   with equality only if $|b| = |a|^{1/(s-1)}$. Taking $a = |X|/E^{1/r}(|X|^r|\mathcal{D})$ and $b = |Y|/E^{1/s}(|Y|^s|\mathcal{D})$, we find that

   $$\frac{|X||Y|}{E^{1/r}(|X|^r|\mathcal{D})E^{1/s}(|Y|^s|\mathcal{D})} \leq \frac{|X|^r}{rE(|X|^r|\mathcal{D})} + \frac{|Y|^s}{sE(|X|^s|\mathcal{D})},$$

   and taking conditional expectations across this inequality and using (16) gives

   $$\frac{E\{|X||Y||\mathcal{D}\}}{E^{1/r}(|X|^r|\mathcal{D})E^{1/s}(|Y|^s|\mathcal{D})} \leq \frac{1}{r} + \frac{1}{s} = 1\,.$$

This yields $E\{|X||Y||\mathcal{D}\} \le E^{1/r}(|X|^r|\mathcal{D})E^{1/s}(|Y|^s|\mathcal{D})$ with equality if and only if

$$\frac{|Y|}{E^{1/s}(|X|^s|\mathcal{D})} = \left(\frac{|X|}{E^{1/r}(|X|^r|\mathcal{D})}\right)^{1/(s-1)} \qquad \text{a.s.}$$

Liapunov's inequality: Suppose that $E|X|^q < \infty$. and let $0 < p \le q$. Then by the conditional Hölder inequality with $1/r = p/q$, $1/s = 1 - p/q$, we find that

$$E(|X|^p|\mathcal{D}) \le E(|X|^q|\mathcal{D})^{p/q}E(1^{1/(1-p/q)}|\mathcal{D})^{1-p/q} = E(|X|^q|\mathcal{D})^{p/q} \qquad \text{a.s.}$$

This implies that $E(|X|^p|\mathcal{D})^{1/p} \le E(|X|^q|\mathcal{D})^{1/q}$ a.s.
Minkowski's inequality: This follows from the conditional Hölder inequality in the same way that Minkowski's inequality follows from the unconditional Hölder inequality.
Jensen's inequality: see the nice proof in Williams, page 89, and note the "important corollary" to Williams' (h).

(26): Suppose that $E(XY) = E(Xh)$ for all $\mathcal{D}-$measurable rv's X. Then, in particular with $h = 1_D$ for $D \in \mathcal{D}$, we have $E(1_D Y) = E(1_D h)$ for $D \in \mathcal{D}$, and hence h is a version (or "determination") of $E(Y|\mathcal{D})$. On the other hand, suppose that h is a version of $E(Y|\mathcal{D})$; i.e. $E(1_D Y) = E(1_D h)$ for all $D \in \mathcal{D}$. Note that this implies $E(1_D Y^+) = E(1_D h^+)$ and $E(1_D Y^-) = E(1_D h^-)$ for all $D \in \mathcal{D}$.

Suppose first that $X \ge 0$. Then there is a sequence of $\mathcal{D}-$measurable simple functions $X_n = \sum_{j=1}^n d_j 1_{D_j} \nearrow X$. Then by the monotone convergence theorem

$$
\begin{aligned}
E(XY) &= E(X(Y^+ - Y^-)) = E(XY^+) - E(XY^-) \\
&= \lim_n E(X_n Y^+) - \lim_n E(X_n Y^-) \qquad \text{by the MCT} \\
&= \lim_n E(\sum_1^n d_j 1_{D_j} Y^+) - \lim_n E(\sum_1^n d_j 1_{D_j} Y^-) \\
&= \lim_n \sum_1^n d_j E(1_{D_j} Y^+) - \lim_n \sum_1^n d_j E(1_{D_j} Y^-) \\
&= \lim_n \sum_1^n d_j E(1_{D_j} h^+) - \lim_n \sum_1^n d_j E(1_{D_j} h^-) \text{by the equality for sets} \\
&= \lim_n E(X_n h^+) - \lim_n E(X_n h^-) \qquad \text{by reversing the above steps} \\
&= E(Xh^+) - E(Xh^-) \qquad \text{by the MCT} \\
&= E(X(h^+ - h^-)) = E(Xh)\,.
\end{aligned}
$$

Now suppose that X is arbitrary with $E|XY| < \infty$. Then

$$
\begin{aligned}
E(XY) &= E((X^+ - X^-)Y) = E(X^+Y) - E(X^-Y) \\
&= E(X^+h) - E(X^-h) \qquad \text{by the result for } X \geq 0 \\
&= E((X^+ - X^-)h) = E(Xh)\,.
\end{aligned}
$$

6. Exercise 7.4.2, part B, page 134, PfS. Redo the calculations in Example 4.1, page 133, but when the sampling is done without replacement. When the sampling is done without replacement the joint probability distribution for (X_1, X_2) is as follows:

			X_1		
		1	2	3	
	1	0	2/30	3/30	5/30
X_2	2	2/30	2/30	6/30	10/30
	3	3/30	6/30	6/30	15/30
		5/30	10/30	15/30	1

Solution: (a) When the sampling is done without replacement the joint probability distribution for (X_1, X_2) is as follows: (this is the table in the problem statement re-arranged to look more like Shorack's table (b), page 132)

		5/30	10/30	15/30	1
	3	3/30	6/30	6/30	15/30
X_2	2	2/30	2/30	6/30	10/30
	1	0	2/30	3/30	5/30
		1	2	3	
			X_1		

Next we tabulate the conditional probabilities along the lines of Shorack's tables (d):

	3	3/8	0	0	
X_2	2	1/2	3/8	0	
	1	0	1/2	3/8	
		1	2	3	
			X_1		

$P(Y = 1|\mathcal{D})(\cdot)$

	3	1/4	1/2	0	
X_2	2	1/2	1/4	1/2	
	1	0	1/2	1/4	
		1	2	3	
			X_1		

$P(Y = 2|\mathcal{D})(\cdot)$

	3	3/8	1/2	1	
X_2	2	0	3/8	1/2	
	1	0	0	3/8	
		1	2	3	
			X_1		

$$P(Y = 3|\mathcal{D})(\cdot)$$

Hence the marginal distribution of $S = X_1 + X_2$ is given by

k	3	4	5	6	
$P(S = k)$	4/30	8/30	12/30	6/30	1

It is easy to compute the conditional distribution of $Y = X_2$ given S (or given $\mathcal{D} = S^{-1}(\mathcal{B})$): letting $D_j = [S = j]$,

		Y		
	1	2	3	$E(Y\|\mathcal{D})$
D_3	1/2	1/2	0	3/2
D_4	3/8	2/8	3/8	2
D_5	0	1/2	1/2	5/2
D_6	0	0	1	3
$P(Y = i)$	5/30	10/30	15/30	

Note that

$$P(Y = i|\mathcal{D}) = \sum_{j=3}^{6} \frac{P([Y = i] \cap D_j)}{P(D_j)} 1_{D_j}$$

satisfies $P(Y = i) = E\{P(Y = i|\mathcal{D})\}$. Also note that $E(Y) = 7/3$, and

$$E(E(Y|\mathcal{D})) = (3/2)(4/30) + 2(8/30) + (5/2)(12/30) + 3(6/30) = 7/3\,.$$

7. 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: (See e.g. Exercise 9.2, Williams, *Probability with Martingales*, page 231.) Suppose first that $X, Y \in L_2(P)$. Then, by Pythagoras (i.e. the orthogonality proved in the solution of problem 1),

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

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

$$E(X^2) = E(Y^2) + E((X - Y)^2)\,. \tag{6}$$

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

$$E(Y^2) = E(X^2) + E((Y-X)^2)\,. \tag{7}$$

Adding (6) and (7) 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

$$\begin{aligned} & E(X-Y)1_{[X>c,Y\le c]} + E(X-Y)1_{[X\le c,Y\le c]} \\ = \; & E(X-Y)1_{[Y\le c]} = E(X1_{[Y\le c]}) - E(Y1_{[Y\le c]}) \\ = \; & E(E(X1_{[Y\le c]}|Y)) - E(Y1_{[Y\le c]}) \\ = \; & E(1_{[Y\le c]}E(X|Y)) - E(Y1_{[Y\le c]}) \\ = \; & E(1_{[Y\le c]}Y) - E(Y1_{[Y\le c]}) = 0 \end{aligned} \tag{8}$$

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

$$\begin{aligned} & E(Y-X)1_{[Y>c,X\le c]} + E(Y-X)1_{[Y\le c,X\le c]} \\ = \; & E(Y-X)1_{[X\le c]} = 0\,. \end{aligned} \tag{9}$$

Adding (8) and (9) yields

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

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 $P(X = Y) = 1$.