Statistics 522, Problem Set 1 Solutions

Wellner; 1/16/2013

1. PfS, exercise 3.2.3, page 42: Consider a measure space $(\Omega, \mathcal{A}, \mu)$. Let $\mu_0 \equiv \mu|_{\mathcal{A}_0}$ for a sub $\sigma-$field $\mathcal{A}_0$ of $\mathcal{A}$. Starting with indicator functions, show that $\int X d\mu = \int X d\mu_0$ for any $\mathcal{A}_0-$measurable function X.

   Remark: Note that this result was used in our proof of the existence of conditional expectation.

   Solution: (a) Suppose first that $X = 1_{D^*}$ where $D^* \in \mathcal{A}_0$. Then since $D^* \in \mathcal{A}_0$

$$\int 1_{D^*} d\mu = \mu(D^*) = \mu_0(D^*) = \int 1_{D^*} d\mu_0.$$

   Thus the claimed identity holds for indicator functions.
   (b) Suppose that $X = \sum_{j=1}^m a_j 1_{D_j}$ for $a_j \in \mathbb{R}$ and $D_j \in \mathcal{A}_0$ for $j = 1, \ldots, m$. Then

$$\begin{aligned}
\int X d\mu &= \int \sum_{j=1}^m a_j 1_{D_j} d\mu = \sum_{j=1}^m a_j \int 1_{D_j} d\mu \\
&= \sum_{j=1}^m a_j \int 1_{D_j} d\mu_0 \quad \text{by part (a)} \\
&= \int \sum_{j=1}^m a_j 1_{D_j} d\mu_0 \quad \text{by linearity of the integral} \\
&= \int X d\mu_0.
\end{aligned}$$

   (c) If $X \geq 0$ is $\mathcal{A}_0-$measurable, then there exist simple functions $X_m \nearrow X$ which are also $\mathcal{A}_0-$measurable. Then, by the monotone convergence theorem,

$$\begin{aligned}
\int X d\mu &= \lim_m \int X_m d\mu = \lim_m \int X_m d\mu_0 \quad \text{by part (b)} \\
&= \int X d\mu_0 \quad \text{by monotone convergence again.}
\end{aligned}$$

(d) If X is a general $\mathcal{A}_0$ measurable function, then write $X = X^+ - X^-$ where X^+ and X^- are non-negative. Then by linearity of the integral and (c) it follows that

$$\begin{aligned}\int X d\mu &= \int (X^+ - X^-) d\mu = \int X^+ d\mu - \int X^- d\mu \\ &= \int X^+ d\mu_0 - \int X^- d\mu_0 \qquad \text{by (c)} \\ &= \int (X^+ - X^-) d\mu_0 = \int X d\mu_0.\end{aligned}$$

2. Exercise 7.4.1, page 131, PfS Course Notes (or Exercise 8.4.1, page 159, PfS (2000): 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

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

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

$$E(Y|\mathcal{D}) = \sum_i \left\{ \frac{1}{P(D_i)} \int_{D_i} Y dP \right\} 1_{D_i}. \tag{2}$$

Solution: We need to show that the quantity on the right side of (1) 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 (1) and $B \in \mathcal{D}$ let

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

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 (1) is a version of $P(A|\mathcal{D})$.

To see that the right side of (2) is a version of $E(Y|\mathcal{D})$ in this case, we need to show that

$$E\{1_D E(Y|\mathcal{D})\} = E\{1_D Y\} \qquad \text{for all } \; D \in \mathcal{D}. \tag{3}$$

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 (3) holds and the proof is complete.

3. Exercise 7.4.2, page 134, PfS Couse Notes with "Discussion 4.2" changed to "Discussion 4.3" (or Exercise 8.4.2, page 161, PfS (2000) plus the following: (C) Mimic discussion 4.2 in case $T = X_1 + X_2$ instead.)
   That is:
   A. (i) Mimic the discussion 4.3 is case $T = X_1 + X_2$ (instead of $T = (X_1^2 + X_2^2)^{1/2}$).
   (ii) Make up another interesting example.

B. (iii) Repeat example 4.1 and the accompanying figure, but now in the context of sampling without replacement.
(iv) Make up another interesting example.

Solution: A. (i) Mimic the discussion 4.3 is case $T = X_1 + X_2$ (instead of $T = (X_1^2 + X_2^2)^{1/2}$).
When $T = T(\underline{X}) = X_1 + X_2$, $T = t$ defines a line

$$L_t \equiv \{(x_1, x_2) : \ x_1 + x_2 = t\}$$

in the plane $\Omega = \mathbb{R}^2$. Let $B \in \mathcal{B}$ denote a one-dimensional Borel set of t's, and then let $D \equiv T^{-1}(B) = \cup_{t \in B} L_t$. Requirements (6) and (7) become, for $A \in \mathcal{B}_2$,

$$\begin{aligned} P(AD) &= P(A \cap \cup_{t\in B} L_t) = \int_{\cup_{t\in B} L_t} P(A|T)(\underline{x}) dP_{\underline{X}}(\underline{x}) \\ &= \int_{\cup_{t\in B} L_t} h_A(\underline{x}) dP_{\underline{X}}(\underline{x}) \\ &= \int_B g_A(t) dP_T(t) = \int_B P(A|T=t) dP_T(t). \end{aligned}$$

So if $g_A(\cdot)$ is given a value at t indicating the probabilistic proportion of $A \cap L_t$ that belongs to A or $h_A(\underline{x})$ is given this same value at all $\underline{x} \in L_t$, then the above equation ought to be satisfied. [When densities exist, such a value would seem to be

$$g_A(t) = \frac{\int_{L_t} 1_A(\underline{x}) p_{\underline{X}}(\underline{x}) d\underline{x}}{\int_{L_t} p_{\underline{X}}(\underline{x}) d\underline{x}},$$

while $h_A(\underline{x}) = g_A(T(\underline{x}))$ would be assigned this same value at each $\underline{x} \in L_t$.] Requirements (1) and (2) become

$$\int_{\cup_{t\in B} L_t} Y dP = \int_{\cup_{t\in B} L_t} h(\underline{x}) dP_{\underline{X}}(\underline{x}) = \int_B g(t) dP_T(t).$$

[When densities exist, then

$$E(Y|T=t) = g(t) = \frac{\int_{L_t} Y(\underline{x}) p(\underline{x}) d\underline{x}}{\int_{L_t} p(\underline{x}) d\underline{x}}$$

seems appropriate with $h(\underline{x})$ getting the same value for all $\underline{x} \in L_t$.]
(ii) Make up another interesting example.

B. (iii) When the sampling is done without replacement the joint probability distribution for (X_1, X_2) is as follows:

			X_1		
		1	2	3	
X_2	1	0	2/30	3/30	5/30
	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

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

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

Here is a version of the tables arranged in roughly the same way as those in PfS Course Notes, page 132:

Table (a): The sum $S = X_1 + X_2$

$Y = X_2$				
3	4	5	6	
2	3	4	5	
1	2	3	4	
	1	2	3	X_1

Table (b): The probabilities $p_{i,j} = P(X_1 = i, X_2 = j)$

$Y = X_2$				
3	3/30	6/30	6/30	
2	2/30	2/30	6/30	
1	0	2/30	3/30	
	1	2	3	X_1

Table (d1): The function $P(Y = 1|\mathcal{D})(\omega)$

$Y = X_2$				
3	3/8	0	0	
2	1/2	3/8	0	
1	any	1/2	3/8	
	1	2	3	X_1

Table (d2): The function $P(Y = 2|\mathcal{D})(\omega)$

$Y = X_2$				
3	1/4	1/2	0	
2	1/2	1/4	1/2	
1	any	1/2	1/4	
	1	2	3	X_1

Table (d3): The function $P(Y = 3|\mathcal{D})(\omega)$

$Y = X_2$				
3	3/8	1/2	1	
2	0	3/8	1/2	
1	any	0	3/8	
	1	2	3	X_1

Table (e): The function $E(Y|\mathcal{D})(\omega)$

$Y = X_2$				
3	2	5/2	3	
2	3/2	2	5/2	
1	any	3/2	2	
	1	2	3	X_1

Table (f): The function $P(Y = y|S = k)$

$Y = X_1$				$E(Y|S = k)$
3	3/8	1/2	1	3
2	1/2	2/8	1/2	5/2
1		1/2	3/8	
	3	4	5	X_2
$E(Y|S = k)$		3/2	2	

4. Exercise 7.4.4, page 139, PfS. (In proving the statement (26), page 177, 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: we did this in class. [But also 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 $X = 1_D$ for $D \in \mathcal{D}$, we have $E(1_DY) = E(1_Dh)$ 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. $E(1_DY) = E(1_Dh)$ for all $D \in \mathcal{D}$. By (20) of Theorem 7.4.1 we have

$$\begin{aligned} E(XY) &= E\{E(XY|\mathcal{D})\} \\ &= E\{XE(Y|\mathcal{D})\} \\ &= E\{Xh\} \end{aligned}$$

for all $\mathcal{D}-$measurable random variables X.

Another route for the part is via the standard machine: Note that $E(1_DY) = E(1_Dh)$ for all $D \in \mathcal{D}$ implies $E(1_DY^+) = E(1_Dh^+)$ and $E(1_DY^-) = E(1_Dh^-)$ 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

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

5. Exercise 8.4.5, page 180, PfS. 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 $g(z) = |z|^r$,

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

   But then by monotonicity of expectation

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

   When $\mu_Y = 0$ this yields $E|X|^r \leq E|X+Y|^r$.