Section 5: Conditional expectations and convergence (solutions)

ARE 210

September 26, 2017

1) **Sum of Bernoulli**: Let $X_{n1}, \ldots, X_{nn}$ be iid Bern$[p_n]$, and let $S_n = \sum_{j=1}^{n} X_{nj}$. Suppose $np_n \to \lambda$. Show that $S_n \xrightarrow{\text{d}} \text{Pois}[\lambda]$.

Hint: the characteristic function of a Pois$[\lambda]$ RV is $\Psi(t) = \exp(\lambda(\exp(it) - 1))$.

First, since X_{nj} are iid, $\Psi_{S_n}(t) = \mathbf{E}[\exp(iX_{nj}t)]^n$. Calculating the expectation yields $\Psi_{S_n}(t) = [1+p_n(\exp(it)-1)]^n$. Taking logs yields $\log \Psi_{S_n}(t) = n \log[1+p_n(\exp(it)-1)]$. Since $np_n \to \lambda$, $p_n \to \frac{\lambda}{n} \to 0$. Therefore $p_n(\exp(it) - 1) \to 0$. Additionally, note that $\lim_{x \to 0} \log(1 + x) = x$. This yields $\log \Psi_{S_n}(t) \to np_n(\exp(it) - 1) \to \lambda(\exp(it) - 1)$. Exponentiating yields $\Psi_{S_n}(t) \to \exp(\lambda(\exp(it)-1))$, which is the characteristic function of a Pois$[\lambda]$ random variable.

2) **Law of iterated expectations**: Show $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y]$.

First, we use that $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[\mathbf{E}[Y|X]|\{\emptyset, \Omega\}]$. By definition of conditional expectation with respect to a σ-algebra, 1) $\mathbf{E}[\mathbf{E}[Y|X]]$ must be constant (in order to be $\{\emptyset, \Omega\}$-measurable), and 2) $\int \mathbf{E}[\mathbf{E}[Y|X]](\omega)\mathbf{1}_\Omega(\omega)P[d\omega] = \int \mathbf{E}[Y|X](\omega)\mathbf{1}_\Omega(\omega)P[d\omega]$. We can simplify the left hand side to $\mathbf{E}[\mathbf{E}[Y|X]]$ since it is constant, and P integrates to 1 over Ω. For the right hand side, substituting in the definition of conditional expectation means it equals $\int Y(\omega)\mathbf{1}_\Omega(\omega)P[d\omega] = \int Y(\omega)P[d\omega] = \mathbf{E}[Y]$. Putting it together yields $\mathbf{E}[\mathbf{E}[Y|X]] = \mathbf{E}[Y]$.

3) **LATE**: Let (Y_1, Y_0, D_1, D_0, Z) be a vector of random variables, where D_1, D_0, and Z are binary. Assume $(Y_1, Y_0, D_1, D_0) \perp Z$ (independence), $\mathbf{E}[D_1] > \mathbf{E}[D_0]$ (first stage), and $D_1 \geq D_0$ almost everywhere (monotonicity). Assume $D = D_1 Z + D_0(1 - Z)$, $Y = Y_1 D + Y_0(1 - D)$, and Z are observed.

a) What is $\mathbf{E}[Y|D = 1] - \mathbf{E}[Y|D = 0]$? What are necessary and sufficient conditions for it to equal $\mathbf{E}[Y_1 - Y_0|D = 1]$?

$$
\begin{aligned}
\mathbf{E}[Y|D=1]-\mathbf{E}[Y|D=0] &= \mathbf{E}[Y_1|D=1]-\mathbf{E}[Y_0|D=0] \\
&= \underbrace{\mathbf{E}[Y_1-Y_0|D=1]}_{\text{TOT}}+\underbrace{\mathbf{E}[Y_0|D=1]-\mathbf{E}[Y_0|D=0]}_{\text{selection}}
\end{aligned}
$$

It equals $\mathbf{E}[Y_1-Y_0|D=1]$ if $\mathbf{E}[Y_0|D=1]-\mathbf{E}[Y_0|D=0]=0$. Note that this is implied by $(Y_1,Y_0)\perp D$

b) Show that $\frac{\mathbf{E}[Y|Z=1]-\mathbf{E}[Y|Z=0]}{\mathbf{E}[D|Z=1]-\mathbf{E}[D|Z=0]}=\mathbf{E}[Y_1-Y_0|D_1>D_0]$.

$$
\begin{aligned}
\mathbf{E}[D|Z=1]-\mathbf{E}[D|Z=0] &= \mathbf{E}[D_1|Z=1]-\mathbf{E}[D_0|Z=0] \\
&= \mathbf{E}[D_1]-\mathbf{E}[D_0] \\
&= \mathbf{E}[D_1-D_0] \\
&= P[D_1>D_0]
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{E}[Y|Z=1]-\mathbf{E}[Y|Z=0] &= \mathbf{E}[D_1Y_1+(1-D_1)Y_0|Z=1]-\mathbf{E}[D_0Y_1+(1-D_0)Y_0|Z=0] \\
&= \mathbf{E}[D_1Y_1+(1-D_1)Y_0]-\mathbf{E}[D_0Y_1+(1-D_0)Y_0] \\
&= \mathbf{E}[(Y_1-Y_0)(D_1-D_0)] \\
&= \mathbf{E}[\mathbf{1}[D_1>D_0](Y_1-Y_0)] \\
&= P[D_1>D_0]\mathbf{E}[Y_1-Y_0|D_1>D_0]
\end{aligned}
$$

c) Assume $D=Z$, what does this reduce to?

$\mathbf{E}[Y|D=1]-\mathbf{E}[Y|D=0]=\mathbf{E}[Y_1-Y_0]$

4) **Distribution of Y_d**: Assume the potential outcomes framework with selection on observables, and assume common support ($\mathbf{E}[D|X]\equiv p(X)\in(0,1)$).

a) Temporarily consider the case where $(Y_1,Y_0)\perp D$. How can the marginal distributions of Y_1 and Y_0 be calculated using (Y,D)? Can the joint distribution be calculated? Intuitively, why not?

$P[Y_1\le y]=\mathbf{E}[\mathbf{1}[Y\le y]|D=1]$. If we only use Y, we can't observe Y_1-Y_0. So we can't tell the difference between high Y_0 correspond to low Y_1 and high Y_0 correspond to high Y_1 without additional assumptions (e.g. what if we observed that $\mathbf{E}[Y_0|X]$ and $\mathbf{E}[Y_1|X]$ had positive covariance? What would we need to assume for that to tell us about the distribution of (Y_0, Y_1)?).

b) Now return to the selection on observables assumption. How can the marginal distributions of Y_1 and Y_0 be calculated using (Y, D, X)?

$P[Y_1 \leq y|X] = \mathbf{E}\left[\frac{D\mathbf{1}[Y\leq y]}{p(X)}\middle| X\right]$, following the proof in the homework. Taking expectations of both sides (and applying the law of iterated expectations) yields $\mathbf{E}[Y_1 \leq y] = \mathbf{E}\left[\frac{D\mathbf{1}[Y\leq y]}{p(X)}\right]$. The marginal distribution of Y_0 can be calculated similarly.