STOR 635 HW 10

by Ameer Qaqish

April 11, 2022

1 Solutions

1. Show that if $\{X_i\}_{i\geq 1}$ is a binary exchangeable sequence, then

(i) With probability one, $X_\infty = \lim_{n\to\infty} \frac{1}{n}(X_1 + \cdots + X_n)$ exists.

(ii) If μ is a measure defined by $\mu(A) = P(X_\infty \in A)$, $A \in \mathcal{B}([0,1])$, then for all n, e_i, $1 \leq i \leq n$,

$$P(X_1 = e_1, \ldots, X_n = e_n) = \int_0^1 x^s (1-x)^{n-s} \mu(dx)$$

where $s = e_1 + \cdots + e_n$.

Proof. (i) We have $\frac{1}{n}(X_1 + \cdots + X_n) = E(X_1 \,|\, \mathcal{E}_n) \to E(X_1 \,|\, \mathcal{E})$ a.s. and in L^1. So we can let $X_\infty = E(X_1 \,|\, \mathcal{E})$.

(ii) Let $\Xi_\infty = \mathcal{L}(X_1 \,|\, \mathcal{E})$. By de Finetti's theorem, $\mathcal{L}(X \,|\, \Xi_\infty) = \Xi_\infty^{\otimes \mathbb{N}}$. We have $\mathcal{L}(X_1 \,|\, \Xi_\infty) = \Xi_\infty$. Since $P(X_1 \in \{0,1\}) = 1$, we get $\Xi_\infty(\{0,1\}) = P(X_1 \in \{0,1\} \,|\, \Xi_\infty) = 1$, i.e. Ξ_∞ is Bernoulli. Thus

$$\begin{aligned}
\Xi_\infty &= \mathrm{Ber}(\Xi_\infty(\{1\})) \\
&= \mathrm{Ber}(\mathcal{L}(X_1 \,|\, \mathcal{E})(\{1\})) \\
&= \mathrm{Ber}(P(X_1 = 1 \,|\, \mathcal{E})) \\
&= \mathrm{Ber}(E(X_1 \,|\, \mathcal{E})) \\
&= \mathrm{Ber}(X_\infty).
\end{aligned}$$

Thus for any $A \in \mathcal{B}(\{0,1\}^{\mathbb{N}})$,

$$\begin{aligned}
P(X \in A) &= E(P(X \in A \,|\, \mathrm{Ber}(X_\infty))) \\
&= \int_0^1 P(X_\infty \in dx)\, P(X \in A \,|\, \mathrm{Ber}(X_\infty) = \mathrm{Ber}(x)) \\
&= \int_0^1 P(X_\infty \in dx)\, \mathrm{Ber}(x)^{\otimes \mathbb{N}}(A).
\end{aligned}$$

When $A = \{e_1\} \times \cdots \times \{e_n\} \times \{0,1\} \times \cdots$, we have

$$\begin{aligned}
\mathrm{Ber}(x)^{\otimes \mathbb{N}}(A) &= \mathrm{Ber}(x)(\{e_1\}) \cdots \mathrm{Ber}(x)(\{e_n\}) \\
&= x^s (1-x)^{n-s},
\end{aligned}$$

where

$$s=e_1+\cdots+e_n.$$

Thus

$$P(X_1=e_1,\ldots,X_n=e_n)=P(X\in A)=\int_0^1 P(X_\infty\in dx)\,x^s\,(1-x)^{n-s}. \qquad \square$$

2. Recall the Polya Urn model. An urn has r red balls and b black balls. At each time, a ball is drawn randomly and two balls of the drawn color are put back in the urn.

(i) Let X_i be one if the i-th draw is a red ball and zero otherwise. Show that

$$P(X_1=e_1,\ldots,X_n=e_n)=\frac{[r(r+1)\cdots(r+s-1)][b(b+1)\cdots(b+n-s-1)]}{(b+r)(b+r+1)\cdots(b+r+n-1)}$$

where $s=e_1+\cdots+e_n$. Hence show that $\{X_i\}$ is an exchangeable sequence.

(ii) Let $\bar{X}_n$ be the proportion of red balls in the urn after n draws. By question 1, $\bar{X}_n$ converges almost surely to $\bar{X}_\infty$. From the above formula, it follows that

$P(\bar{X}_n=k\,/\,n)=\binom{n}{k}\dfrac{[r(r+1)\cdots(r+k-1)][b(b+1)\cdots(b+n-k-1)]}{(b+r)(b+r+1)\cdots(b+r+n-1)}$, $n\geq 1$, $0\leq k\leq n$.

From this and characteristic functions (see note PolUrn in resources), one can show that $\bar{X}_\infty$ has a Beta(r,b) distribution (you can assume this without proof for the homework). Show that, taking μ to be the Beta(r,b) distribution, the equality in part (ii) of question 1 indeed holds.

Proof. (i) Let $n\in\mathbb{N}$. For $0\leq i\leq n$, let $s_i=e_1+\cdots+e_i$. We have

$$\begin{aligned}P(X_1=e_1,\ldots,X_n=e_n) &= P(X_1=e_1)\,P(X_2=e_2\,|\,X_1=e_1,X_2=e_2)\cdots P(X_n=e_n\,|\\ &\quad X_1=e_1,\ldots,X_{n-1}=e_{n-1})\\ &= \prod_{1\leq i\leq n:e_i=1}\frac{r+s_{i-1}}{b+r+i-1}\prod_{1\leq i\leq n:e_i=0}\frac{b+i-1-s_{i-1}}{b+r+i-1}\end{aligned}$$

$$=\frac{1}{(b+r)\cdots(b+r+n-1)}\prod_{1\leq i\leq n:e_i=1}(r+s_{i-1})\prod_{1\leq i\leq n:e_i=0}(b+i-1-s_{i-1}).$$

Since the urn has r red balls before drawing the first red ball and has $r+s_n-1$ red balls before drawing the last red ball, we have

$$\prod_{1\leq i\leq n:e_i=1}(r+s_{i-1})=r\,(r+1)\cdots(r+s_n-1).$$

Similarly,

$$\prod_{1\leq i\leq n:e_i=0}(b+i-1-s_{i-1})=b\,(b+1)\cdots(b+n-s_n-1).$$

Since $s_n = s$,

$$P(X_1 = e_1, \ldots, X_n = e_n) = \frac{r\,(r+1)\cdots(r+s-1)\,b\,(b+1)\cdots(b+n-s-1)}{(b+r)\,(b+r+1)\cdots(b+r+n-1)}.$$

This shows that $P(X_1 = e_1, \ldots, X_n = e_n)$ is a function of $e_1 + \cdots + e_n$. Since summation is invariant under permutations, $(X_i)_{1 \le i \le n}$ is exhangeable. Since n was arbitrary, $(X_i)_{i \ge 1}$ is exhangeable.

(ii) Let $\mu = \mathrm{Beta}(r, b)$. This means that μ has density

$$f(x) = \frac{1}{B(r,b)}\, x^{r-1}\,(1-x)^{b-1}, \quad x \in [0,1],$$

where

$$B(r,b) = \frac{\Gamma(r)\,\Gamma(b)}{\Gamma(r+b)}.$$

Let $e_1, \ldots, e_n \in \{0,1\}$ be arbitrary. Let $s = e_1 + \cdots + e_n$. We compute

$$\begin{aligned}\int_0^1 x^s\,(1-x)^{n-s}\,\mu(dx) &= \frac{1}{B(r,b)} \int_0^1 x^{r+s-1}\,(1-x)^{b+n-s-1}\,dx \\ &= \frac{B(r+s, b+n-s)}{B(r,b)} \\ &= \frac{\Gamma(r+b)}{\Gamma(r)\,\Gamma(b)}\,\frac{\Gamma(r+s)\,\Gamma(b+n-s)}{\Gamma(r+b+n)}.\end{aligned}$$

On the other hand,

$$\frac{r\,(r+1)\cdots(r+s-1)\,b\,(b+1)\cdots(b+n-s-1)}{(b+r)\,(b+r+1)\cdots(b+r+n-1)} =$$

$$\frac{(b+r-1)!}{(b+r+n-1)!}\,\frac{(r+s-1)!}{(r-1)!}\,\frac{(b+n-s-1)!}{(b-1)!}.$$

Using the identity $\Gamma(k) = (k-1)!$, it is seen that this expression is equal to $\frac{\Gamma(r+b)}{\Gamma(r)\,\Gamma(b)}\,\frac{\Gamma(r+s)\,\Gamma(b+n-s)}{\Gamma(r+b+n)}$. Thus the equality

$$\frac{r\,(r+1)\cdots(r+s-1)\,b\,(b+1)\cdots(b+n-s-1)}{(b+r)\,(b+r+1)\cdots(b+r+n-1)} = \int_0^1 x^s\,(1-x)^{n-s}\,\mu(dx)$$

indeed holds. $\square$

3. Let $X_1, X_2, \ldots$ be $N(0,1)$ random variables (defined on the same probability space) such that for any $n \in \mathbb{N}$, $(X_1, X_2, \ldots, X_n)$ is jointly Gaussian with $\mathrm{Cov}(X_i, X_j) = \rho$ for $1 \le i < j \le n$. Show that this is indeed an exchangeable family.

(i) Show that the cumulative empirical distribution defined by $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x)$, $x \in \mathbb{R}$, has the following random limit for each $x \in \mathbb{R}$:

$$F_n(x) \xrightarrow{a.s.} F_\infty(x) := \Phi\!\left(\frac{x - Y}{\sqrt{1-\rho}}\right),$$

where Φ is the standard normal cdf and $Y \sim N(0,\rho)$ is the almost sure limit of $\frac{1}{n}(X_1+\cdots+X_n)$ as $n\to\infty$. This observation can be used to show that the $\sigma(\mathcal{F}_\infty)=\sigma(Y)$ (no need to prove this).

(ii) Show that, conditional on $\sigma(Y)$, $\{X_i\}_{i\in\mathbb{N}}$ are iid $N(Y,1-\rho)$.

Proof. For any $n\in\mathbb{N}$ and distinct $i_1,\ldots,i_n\in\mathbb{N}$, $(X_{i_1},\ldots,X_{i_n})$ are jointly Gaussian with mean vector 0 and covariance matrix $\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$. Since the distribution of $(X_{i_1},\ldots,X_{i_n})$ is the same for all $n\in\mathbb{N}$ and all distinct n-tuples $(i_1,\ldots,i_n)$, it follows that $(X_n)_{n\in\mathbb{N}}$ is exchangeable.

We prove part (ii) first.

(ii) Let $k\in\mathbb{N}$ be arbitrary. We will compute the conditional distribution of $(X_1,\ldots,X_k)$ given Y. Let $n\geq k$ be arbitrary. Define

$$\bar{X}_n=\frac{1}{n}(X_1+\cdots+X_n).$$

Note that $(X_1,\ldots,X_k,\bar{X}_n)$ is jointly Gaussian with mean

$$\mu_n=0,$$

and covariance matrix

$$\Sigma_n=\begin{pmatrix} 1 & \rho & \rho & \cdots & \rho & \rho_n \\ \rho & 1 & \rho & \cdots & \rho & \rho_n \\ \rho & \rho & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 1 & \rho & \rho_n \\ \rho & \rho & \cdots & \rho & 1 & \rho_n \\ \rho_n & \rho_n & \cdots & \rho_n & \rho_n & \rho_n \end{pmatrix},$$

where

$$\rho_n=\operatorname{Cov}(X_1,\bar{X}_n)=\frac{1}{n}(1+(n-1)\,\rho).$$

By the backward martingale convergence theorem, $\bar{X}_n=E(X_1\,|\,\mathcal{E}_n)\to E(X_1\,|\,\mathcal{E})=Y$ as $n\to\infty$ a.s. and in L^1. Thus $(X_1,\ldots,X_k,\bar{X}_n)\to(X_1,\ldots,X_k,Y)$ as $n\to\infty$ a.s. and in L^1. Since $\rho_n\to\rho$ as $n\to\infty$, we have

$$\Sigma:=\lim_{n\to\infty}\Sigma_n=\begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & \rho \\ \rho & \rho & \cdots & \rho & \rho \end{pmatrix}.$$

The characteristic function of $(X_1,\ldots,X_k,\bar{X}_n)$ is

$$\varphi_n(t)=\exp\!\left(i\,\mu_n^T t-\frac{1}{2}t^T\Sigma_n t\right)\to\exp\!\left(-\frac{1}{2}t^T\Sigma\,t\right)\text{ as } n\to\infty.$$

Since $(X_1, \ldots, X_k, \bar{X}_n)$ converge to $(X_1, \ldots, X_k, Y)$ a.s., it follows from the DCT that the characteristic function of $(X_1, \ldots, X_k, Y)$ is

$$\varphi(t) = \lim_{n \to \infty} \varphi_n(t) = \exp\left(-\frac{1}{2} t^T \Sigma\, t\right).$$

Thus $(X_1, \ldots, X_k, Y)$ is jointly Gaussian with mean 0 and covariance matrix Σ. Fix distinct $i, j \in \{1, \ldots, k\}$. The covariance matrix of (X_i, X_j, Y) is

$$\Sigma^{i,j,Y} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & \rho \end{pmatrix}.$$

Using the general formula for the conditional distribution of one component of a joint Gaussian conditioned on another, it follows that for $y \in \mathbb{R}$, the conditional distribution of (X_i, X_j) given $Y = y$ is jointly Gaussian with mean vector

$$\mu_y = \begin{pmatrix} \rho \\ \rho \end{pmatrix} \rho^{-1} y = \begin{pmatrix} y \\ y \end{pmatrix}$$

and covariance matrix

$$\begin{aligned} \Sigma_y &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho \\ \rho \end{pmatrix} \rho^{-1} \begin{pmatrix} \rho & \rho \end{pmatrix} \\ &= \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} - \begin{pmatrix} \rho & \rho \\ \rho & \rho \end{pmatrix} \\ &= \begin{pmatrix} 1-\rho & 0 \\ 0 & 1-\rho \end{pmatrix}. \end{aligned}$$

See the answer by Ben at [1] for a proof of the general formula. Thus the conditional distribution of $(X_1, \ldots, X_k)$ given $Y = y$ is jointly Gaussian with mean vector $\begin{pmatrix} y \\ y \\ \vdots \\ y \end{pmatrix}$ and covariance matrix $(1-\rho)\, I$. In other words, given $Y = y$, $X_1, \ldots, X_k$ are i.i.d. $N(y, 1-\rho)$. Since $k \in \mathbb{N}$ was arbitrary, this means that conditional on Y, $(X_i)_{i \in \mathbb{N}}$ are i.i.d. $N(Y, 1-\rho)$.

(i) Let $x \in \mathbb{R}$. Note that $F_n(x) = P(X_1 \leq x \,|\, \mathcal{E}_n)$. By the theorem proved in class, $F_n(x) \to P(X_1 \leq x \,|\, \mathcal{E}) = P(X_1 \leq x \,|\, \mathcal{T})$ as $n \to \infty$ a.s. and in L^1, where $\mathcal{T}$ is the tail σ-algebra of $(X_i)_{i \in \mathbb{N}}$. Thus $F_\infty(x) = P(X_1 \leq x \,|\, \mathcal{T})$. We claim that $P(X_1 \leq x \,|\, \mathcal{T}) = P(X_1 \leq x \,|\, Y)$. Define $f: \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ by $f(y) = I(y_1 \leq x)$. Let Z be an arbitrary $\mathcal{T}$-measurable random variable. For each $y \in \mathbb{R}$, since $(X_i)_{i \in \mathbb{N}}$ are i.i.d. given $Y = y$ and Z is $\mathcal{T}$-measurable, it follows by Kolmogorov's 0-1 law that $Z = E(Z \,|\, Y = y)$ $P(\cdot \,|\, Y = y)$ almost surely and hence

$$\begin{aligned} E(Z f(X) \,|\, Y = y) &= E(E(Z \,|\, Y = y)\, f(X) \,|\, Y = y) \\ &= E(Z \,|\, Y = y)\, E(f(X) \,|\, Y = y). \end{aligned}$$

Thus

$$E(Z\,f(X)\,|\,Y) = E(Z\,|\,Y)\,E(f(X)\,|\,Y).$$

Also,

$$E(Z\,E(f(X)\,|\,Y)\,|\,Y) = E(Z\,|\,Y)\,E(f(X)\,|\,Y).$$

By the tower property,

$$E(Z\,f(X)) = E(Z\,E(f(X)\,|\,Y)).$$

Since Z was an arbitrary bounded $\mathcal{T}$-measurable random variable, it follows that

$$E(f(X)\,|\,\mathcal{T}) = E(E(f(X)\,|\,Y)\,|\,\mathcal{T}).$$

Since Y is $\mathcal{T}$-measurable, $\sigma(Y) \subset \mathcal{T}$. Thus $E(E(f(X)\,|\,Y)\,|\,\mathcal{T}) = E(f(X)\,|\,Y)$, and hence

$$E(f(X)\,|\,\mathcal{T}) = E(f(X)\,|\,Y).$$

So by part (ii),

$$\begin{aligned} P(X_1 \le x\,|\,\mathcal{T}) &= P(X_1 \le x\,|\,Y) \\ &= \Phi\left(\frac{x - Y}{\sqrt{1-\rho}}\right). \end{aligned}$$

Since $F_\infty(x) = P(X_1 \le x\,|\,\mathcal{T})$, this finishes the proof. $\square$

4. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a time homogeneous E-valued Markov chain with distributions $\{P_x\}_{x\in E}$ on $(\Omega, \mathcal{F})$. Let $K: E \times \mathcal{B}(E)^{\otimes\infty} \to [0,1]$ defined as

$$K(x, B) = P_x((X_0, X_1, \dots) \in B), \quad (x, B) \in E \times \mathcal{B}(E)^{\otimes\infty}.$$

Show that for any bounded measurable function $F: E^\infty \to \mathbb{R}$ and $m \in \mathbb{N}_0$,

$$E_x(F(\Theta_m\,\underline{X})\,|\,\mathcal{F}_m) = \int F(y)\,K(X_m, dy), \text{ a.s.},$$

where $\Theta_m\,\underline{X} = (X_m, X_{m+1}, \dots)$.

Proof. We use the notation

$$E_x(F(X)) = \int_\Omega F(X(\omega))\,P_x(d\omega) = \int_{E^\infty} F(y)\,K(x, dy).$$

Thus we must show that

$$E_x(F(\Theta_m\,X)\,|\,\mathcal{F}_m) = E_{X_m}(F(X)).$$

First consider the case where F depends only on the coordinates $x_0, \dots, x_n$ for some $n \ge 0$. Write $F(x) = F(x_0, \dots, x_n)$. Thus $F(\Theta_m\,X) = F(X_m, X_{m+1}, \dots, X_{m+n})$. Let $g: E^n \to \mathbb{R}$ be an arbitrary bounded $\mathcal{B}(E^m)$-measurable function. Let $\kappa: E \times \mathcal{B}(E) \to [0,1]$ be the one step transition probability kernel defined by

$$\kappa(x, A) = P_x(X_1 \in A).$$

We have

$$\begin{aligned}
E_x(g(X_0,\ldots,X_m)\,F(\Theta_m X)) &= E_x(g(X_0,\ldots,X_m)\,F(X_m,\ldots,X_{m+n}))\\
&= \int_E \kappa(x_0,\, d\,x_1)\ \cdots\ \int_E \kappa(x_{m+n-1},\, d\,x_{m+n})\ g(x_1,\,\ldots,\\
&\quad x_m)\,F(x_m,\ldots,x_{m+n})\\
&= \int_E \kappa(x_0, dx_1)\cdots\int_E \kappa(x_{m-1}, dx_m)\, g(x_0,\ldots,x_m)\int_E \kappa(x_m,\\
&\quad dx_{m+1})\cdots\int_E \kappa(x_{m+n-1}, dx_{m+n})\,F(x_m,\ldots,x_{m+n})\\
&= \int_E \kappa(x_0,\ d\,x_1)\ \cdots\ \int_E \kappa(x_{m-1},\ d\,x_m)\ g(x_0,\ \ldots,\\
&\quad x_m)\,E_{x_m}(F(X_0,\ldots,X_n))\\
&= E_x(g(X_0,\ldots,X_m)\,E_{X_m}(F(X_0,\ldots,X_n)))\\
&= E_x(g(X_0,\ldots,X_m)\,E_{X_m}(F(X))).
\end{aligned}$$

Since $g: E^n \to \mathbb{R}$ was an arbitrary bounded $\mathcal{B}(E^m)$-measurable function, this implies that

$$E_x(F(\Theta_m X)\,|\,\mathcal{F}_m) = E_{X_m}(F(X)).$$

Thus the problem is solved for F that depend on finitely many coordinates.

Now suppose $F: E^\infty \to \mathbb{R}$ is bounded and measurable but does not necessarily depend on only finitley many coordinates. By linearity, we may assume that $F \geq 0$. By monotone convergence, we may assume that F is a simple function. By linearity, we may assume that $F = 1_A$ for some $A \in \mathcal{B}(E^\infty)$. The collection $\mathcal{D}$ of sets $A \in \mathcal{B}(E^\infty)$ for which $E_x(1_A(\Theta_m \underline{X})\,|\,\mathcal{F}_m) = E_{X_m}(1_A(X))$ is a λ-system, and, by what was already proved, it contains all sets of the form $A_1 \times \cdots \times A_n \times E \times \cdots$ with $n \in \mathbb{N}$ and $A_1, \ldots, A_n \in \mathcal{B}(E)$. By the π-λ theorem, $\mathcal{D} = \mathcal{B}(E^\infty)$. This completes the proof. □

Bibliography

[1] Deriving the conditional distributions of a multivariate normal distribution.