Let $\theta > 0$ be a constant and let $X_1,\ldots,X_n$ be independent random variables, each uniformly distributed on the interval $[0,\theta]$. Let $Y_n = \prod_{i=1}^{n} X_i$, $Z_n = X_n - \frac{\theta}{2}$, $n \geq 1$.

(a) Show that $\lim_{n\to\infty} Y_n^{1/n} = \frac{\theta}{e}$ a.s.

Assume from now on that $\theta = 2$.

(b) Show that $Y_1, Y_2, \ldots$ is a martingale.

(c) Show that Y_n converges with probability 1 and identify the distribution of the limit random variable Y_∞.

(d) Identify the limit distribution for each of the following 5 sequences and justify your answers.

(i) $R_n = (Z_1 + Z_2 + \ldots + Z_n)/\sqrt{n}$
(ii) $T_n = (Z_1^2 + \ldots + Z_n^2)/n$
(iii) $U_n = (Z_1 + Z_2 + \ldots + Z_n)^2/n$
(iv) $V_n = (Z_1 + \ldots + Z_n)/\sqrt{Z_1^2 + \ldots + Z_n^2}$
(v) $W_n = \sqrt{n}\log(3T_n)$

(e) Show that there do NOT exist iid random variables S_1 and S_2 such that $S_1 - S_2$ has the same distribution as Z_1.

Solution

(a). Consider $H_n = \log(Y_n^{\frac{1}{n}}) = \frac{1}{n}\sum_{i=1}^{n} \log X_i$.

By SLLN, $H_n \to E\log X_1$ a.s.

$E\log X_1 = \log\theta - 1$. It follows that $Y_n^{\frac{1}{n}} = e^{H_n} \to \frac{\theta}{e}$ a.s.

(b). Clearly $E|Y_n| < \infty$ for $n \geq 1$.

$E(Y_{n+1}|Y_n) = E(X_{n+1}\cdot Y_n | Y_n) = Y_n\, E(X_{n+1}|Y_n)$

$= Y_n\, E(X_{n+1})$, where the last equality follows from independence between X_{n+1} and $Y_n = X_1 \cdots X_n$.

For $\theta = 2$, $E X_{n+1} = 1$. Thus $E(Y_{n+1}|Y_n) = Y_n$.

So $Y_1, Y_2, \ldots$ is a martingale.

(c). From (a), $Y_n = (Y_n^{\frac{1}{n}})^n \to 0$ a.s.

(d). $EZ_n = 0$, $EZ_n^2 = \frac{1}{3}$, $Var(Z_n^2) = \frac{1}{5} - \frac{1}{9}$.

(i) By CLT, $R_n \xrightarrow{\mathcal{D}} N(0, \frac{1}{3})$

(ii) By WLLN, $T_n \xrightarrow{\mathcal{D}} \frac{1}{3}$

(iii) From (i), $U_n = R_n^2 \xrightarrow{\mathcal{D}} \frac{1}{3}\chi_1^2$

(iv) $V_n = \frac{Z_1 + \cdots + Z_n}{\sqrt{n}} \Big/ \sqrt{\frac{Z_1^2 + \cdots + Z_n^2}{n}} \xrightarrow{\mathcal{D}} \frac{N(0,\frac{1}{3})}{\sqrt{\frac{1}{3}}} = N(0,1)$.

by Slutsky

(v). Note $\sqrt{n}\,(3T_n - 1)$

$$= 3\sqrt{n}\left(T_n - \tfrac{1}{3}\right)$$

$$= 3 \cdot \frac{(Z_1^2 - \frac{1}{3}) + \cdots + (Z_n^2 - \frac{1}{3})}{\sqrt{n}} \xrightarrow[\text{by CLT}]{\mathcal{D}} N(0, 9\,Var(Z_1^2)) = N(0, \tfrac{4}{5})$$

Applying δ-method, we have that

$$\sqrt{n}\,\log(3T_n) = \sqrt{n}\,(\log(3T_n) - \log 1)$$

$$\xrightarrow{\mathcal{D}} N(0, \tfrac{4}{5}).$$

(e). Assume if possible $S_1 - S_2 = Z_1$ for some i.i.d r.v.'s S_1 and S_2. Let $\varphi(t)$ be the characteristic function of S_1. It is straightforward to find the characteristic function of Z_1, which is $\frac{\sin t}{t}$. Then we must have

$$|\varphi(t)|^2 = \frac{\sin t}{t},$$

which is impossible since $\frac{\sin t}{t}$ can be negative.

Let $(\Omega,\Im,P)$ be a probability space and let $U_1,\ldots,U_n$ be independent random variables with uniform distributions over $(0,1)$. Let $U_{1:n} \le U_{2:n} \le \cdots \le U_{n:n}$ be the corresponding order statistics, and let

$$F_n(t,\omega) := \frac{1}{n}\sum_{i=1}^{n} I_{(0,t]}(U_i(\omega)), \quad 0<t<1$$

be the uniform empirical distribution function.

(a) Show that for fixed $t \in (0,1)$, the distribution of the random variable $nF_n(t,\cdot)$ has a binomial distribution.

(b) Prove that for any $\alpha > 0$ and $k = 0,1,2,\ldots$

$$\lim_{n\to\infty} P\left\{\omega : F_n\left(\frac{\alpha}{n},\omega\right) = \frac{k}{n}\right\} = e^{-\alpha}\alpha^k / k!$$

(c) Suppose that $p_n \to p \in (0,1)$. Prove that

$$\sup_{-\infty<x<\infty}\left|P\left\{\omega : F_n(p_n,\omega) - p_n < \frac{x}{\sqrt{n}}\right\} - \Phi\left(\frac{x}{\sqrt{p(1-p)}}\right)\right| \to 0, \quad \text{as } n\to\infty,$$

where Φ is the cumulative distribution function of the standard normal distribution.

(d) Show that for $1 \le i \le n$,

$$(*) \qquad P\{\omega : U_{i:n}(\omega) \le t\} = P\left\{\omega : F_n(t,\omega) \ge \frac{i}{n}\right\}, \quad 0<t<1.$$

Infer that $U_{i:n}$ has a continuously differentiable cumulative distribution function, and determine its Lebesgue density.

(e) Use (*) and (c) to prove that for any $p \in (0,1)$,

$$\sqrt{n}\left(U_{\lfloor np \rfloor:n} - p\right)$$

converges in distribution to a normal random variable with mean zero and variance $p(1-p)$, where $\lfloor np \rfloor$ denotes the integer part of np.

Solution

(a) Note that $\sum_{i=1}^{n} I_{(0,t]}(U_i(\omega))$ is the sum of iid Bernoulli r.v.'s $X_i(\omega) = I_{(0,t]}(U_i(\omega))$. So $n F_n(t, \cdot)$ has $Bin(n, p)$ distribution, where

$p = P(X_i = 1) = P(0 < U_i \le t) = t.$

(b) $P(\omega : F_n(\frac{a}{n}, \omega) = \frac{k}{n})$

$= P(\omega : \sum_{i=1}^{n} X_i = k)$

$\overset{\text{from (a)}}{=} \binom{n}{k} \left(\frac{a}{n}\right)^k \left(1 - \frac{a}{n}\right)^{n-k} \longrightarrow \frac{e^{-\alpha} \cdot a^k}{k!}$ for each $k = 0$,

This is Poisson approximation of $Bin(n, p)$ when $np \to \alpha$.

(c). Let $X_{n,i} = I_{(0, p_n]}(U_i) - p_n$, $i = 1, 2, \cdots, n$.
It forms a triangular array.
Note that $E X_{n,i} = 0$, $Var(X_{n,i}) = p_n(1 - p_n)$.
$S_n^2 = \sum_{i=1}^{n} Var(X_{n,i}) = n p_n (1 - p_n) \to np(1-p) > 0$.
It is straightforward to verify that Lindeberg condition holds. As a consequence,

$$\frac{\sum_{i=1}^{n} X_{n,i}}{\sqrt{n p_n (1 - p_n)}} \xrightarrow{\mathcal{D}} N(0,1),$$

i.e., $\sqrt{n}\,(F_n(p_n, \cdot) - p_n) \xrightarrow{\mathcal{D}} N(0, p(1-p))$

Since normal distribution is continuous, by Polya Theorem, the conclusion of (c) follows.

(d). $U_{i:n}(\omega) \le t \iff$ at least i of $U_1(\omega), \dots, U_n(\omega)$ are no bigger than $t \iff F_n(t,\omega) \ge \frac{i}{n}$. So (*) follows. Then

$$P(U_{i:n} \le t) = \sum_{j=i}^{n} \binom{n}{j} t^j (1-t)^{n-j} \qquad 0 \le t \le 1.$$

The density function is

$$\sum_{j=i}^{n} \binom{n}{j} j t^{j-1} (1-t)^{n-j} - \sum_{j=i}^{n} \binom{n}{j} t^j (n-j)(1-t)^{n-j-1}$$

$$= \sum_{j=i}^{n} \frac{j \cdot n!}{j!(n-j)!} t^{j-1}(1-t)^{n-j} - \sum_{j=i}^{n-1} \frac{(n-j)\cdot n!}{j!(n-j)!} t^{j}(1-t)^{n-j-1}$$

$$= \sum_{j=i}^{n} \frac{n!}{(j-1)!(n-j)!} t^{j-1}(1-t)^{n-j} - \sum_{\tilde{j}=i+1}^{n} \frac{n!}{(\tilde{j}-1)!(n-\tilde{j})!} t^{\tilde{j}-1}(1-t)^{n-\tilde{j}}$$

$$= \frac{n!}{(i-1)!(n-i)!} t^{i-1}(1-t)^{n-i}$$

It is a Beta$(i, n-i+1)$ distribution.

(e). $\sqrt{n}\,(U_{\lfloor np \rfloor:n} - p) \le x$

$\iff U_{\lfloor np \rfloor:n} \le \frac{x}{\sqrt{n}} + p$

$\iff F_n(p + \frac{x}{\sqrt{n}}, \cdot) \ge \frac{\lfloor np \rfloor}{n}$ (by (*))

Thus

$$P(\sqrt{n}\,(U_{\lfloor np\rfloor:n} - p) \le x)$$

$$= P\left(F_n(p+\tfrac{x}{\sqrt{n}},\cdot) - (p+\tfrac{x}{\sqrt{n}}) \geq \frac{\lfloor np\rfloor}{n} - p - \frac{x}{\sqrt{n}} \right)$$

$$= \Bigg[P\left(F_n(p+\tfrac{x}{\sqrt{n}},\cdot) - (p+\tfrac{x}{\sqrt{n}}) \geq \frac{-x-\sqrt{n}(p-\frac{\lfloor np\rfloor}{n})}{\sqrt{n}} \right)$$

$$- \left(1 - \Phi\left(\frac{-x-\sqrt{n}(p-\frac{\lfloor np\rfloor}{n})}{\sqrt{p(1-p)}}\right)\right) \Bigg]$$

$$+ \left\{ \left(1-\Phi\left(\frac{-x-\sqrt{n}(p-\frac{\lfloor np\rfloor}{n})}{\sqrt{p(1-p)}}\right)\right) - \left(1-\Phi\left(\frac{-x}{\sqrt{p(1-p)}}\right)\right) \right\}$$

$$+ \; 1 - \Phi\left(\frac{-x}{\sqrt{p(1-p)}}\right)$$

The quantity in [ ] goes to 0 from (c), and the quantity in { } goes to 0 because Φ is continuous and $\sqrt{n}\,(p - \frac{\lfloor np\rfloor}{n}) \to 0$.

It follows that

$$P(\sqrt{n}\,(U_{\lfloor np\rfloor:n} - p) \le x) \to 1 - \Phi\left(\frac{-x}{\sqrt{p(1-p)}}\right) = \Phi\left(\frac{x}{\sqrt{p(1-p)}}\right),$$

i.e.,

$$\sqrt{n}\,(U_{\lfloor np\rfloor:n} - p) \xrightarrow{\mathcal{D}} N(0,\, p(1-p)).$$

Ph.D. Prelims **Theory III** **Spring 2000**

Let $(X_1, X_2, \ldots, X_k)$, $k \geq 3$, be a multinomial $(n; p_1, p_2, \ldots, p_k)$ random vector with n and k known and $p_i > 0$, $\sum_{i=1}^{k} p_i = 1$ unknown.

a) Justify your answers to the following questions:
   1) Is this a dominated family?
   2) Is this an exponential family?
   3) Is this family identifiable?
   4) Is there a minimal sufficient statistic?

b) What are all functions $\gamma(p_1, p_2, \ldots, p_k)$ that admit unbiased estimators?

c) Let $r_1, r_2, \ldots, r_k$ be nonnegative integers such that $\sum_{i=1}^{k} r_i = n$. Find the UMVU estimator of $p_1^{r_1}\ p_2^{r_2} \ldots p_k^{r_k}$.

d) Now suppose that $n \to \infty$. Find a consistent estimator of p_2/p_1. Is it asymptotically normal? What is its asymptotic variance?

e) For the above statistical experiment define the parameter space Θ and the family $\{P_\theta, \theta \in \Theta\}$. Find a subset A of the parameter space Θ such that the statistic $(X_1+X_2, X_3, \ldots, X_k)$ is sufficient for $\{P_\theta, \theta \in A\}$. (Hint: Look at the conditional distribution of (X_1, X_2) given X_1+X_2).

f) Suppose $\theta = (p_1, p_2, \ldots, p_k)$ has a prior distribution with a p.d.f. proportional to $p_1^{\alpha_1 - 1}\ p_2^{\alpha_2 - 1} \ldots p_k^{\alpha_k - 1}$ with $\alpha_i > 0$ for all i.

   (i) Find the posterior distribution of θ, given the data $(X_1, X_2, \ldots, X_k)$ (the constant of integration need not be evaluated).

   (ii) Find the Bayes estimator of p_i under squared error loss.

Solution to Ph D. Prelims (2000) Theory III

(a) 1. Yes, by the counting measure on $S \equiv \{(x_1, x_2, \ldots x_k);$
$x_i \geq 0, \quad \sum_1^k x_i = n, \quad x_i$ nonnegative integer$\}$

2. Yes, since the probability $(X_1 = x_1, \quad X_k = x_k)$ can be written as

$$\frac{n!}{x_1! \, x_2! \quad x_k!} \exp\left\{ \sum_1^k x_i \ln p_i \right\}$$

3. Yes, since $p = (p_1, \ldots p_k) \neq p' = (p_1', \ldots, p_k')$ implies that $\exists \; (x_1, x_2, \ldots x_k)$ such that
$$P_p(x_1, x_2, \quad x_k) \neq P_{p'}(x_1, x_2, \ldots x_k)$$

4. Yes, $(X_1, X_2, \quad X_k)$ is minimal suff. by the Halmos Savage theorem

b) If $h(X_1, X_2, \quad X_k)$ is an unbiased estimator of $\gamma(p_1, p_2, \ldots p_k)$ then

$$\gamma(p_1, p_2, \quad p_k) = E_p h(X_1, \quad X_k)$$
$$= \sum_{(x_1, \, x_k) \in S} \frac{n!}{x_1! \, x_2! \quad x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} h(x_1, x_2, \ldots x_k)$$

Thus it is n d s that γ be a multinomial in $p_1, p_2, \quad p_k$ of degree $\leq n$.

c) Set $h(x_1, x_2, \quad x_k) = \dfrac{x_1! \, x_2! \quad x_k!}{n!}$ if $(x_1, \quad x_k) = (x_1, x_2, \ldots x_k)$ and 0 otherwise.

Then $E_p h(X_1, \quad X_k) = p_1^{x_1} p_2^{x_2} \quad p_k^{x_k}$

By the completeness and sufficiency of $X_1, X_2, \quad X_k$ the Lehmann Scheffe Theorem asserts that this is the UMVU.

d) A strongly consistent estimator of p_1 is $\frac{X_1}{n}$, p_2 is $\frac{X_2}{n}$ and so a strongly consistent estimator of p_2/p_1 is $\dfrac{X_2}{X_1}$

By the delta method applied to $f(x_1, x_2) = \frac{x_2}{x_1}$

$\frac{X_2}{X_1}$ is asymptotically normal

More precisely $\sqrt{n}\left(\frac{X_2}{X_1} - \frac{p_2}{p_1} \right) \xrightarrow{d}$

$$N\left(0,\ \left(-\frac{p_2}{p_1^2}\right)^2 p_1(1-p_1) + \frac{1}{p_1^2} p_2(1-p_2) + 2\frac{p_2}{p_1^3} p_1 p_2 \right)$$

e) The conditional distribution of (X_1, X_2) given $X_1 + X_2$ is Binomial with $\left(X_1+X_2, \frac{p_1}{p_1+p_2} \right)$

So a n + s condition for sufficiency of $(X_1+X_2, X_3, \ldots X_k)$ is that $\frac{p_1}{p_1+p_2}$ be constant on A.

f) The posterior has a density proportional to

$$p_1^{\alpha_1+x_1-1} \quad p_2^{\alpha_2+x_2-1} \quad \cdots \quad p_k^{\alpha_k+x_k-1}$$

The Bayes estimator of p_1 under squared error loss is the (conditional) expectation of p_1 w.r.t. the above posterior

Ph.D. Prelims **Theory IV** **Spring 2000**

Let (X,Y,Z) be a random vector with $EX = EY = EZ = 0$, $E(X^2 + Y^2 + Z^2) < \infty$ and covariance matrix Σ.

a) Find $\hat{X}_z$ and $\hat{Y}_z$, the best (in the least squares sense) linear unbiased predictors of X and Y respectively given Z.

b) Let $U = X - \hat{X}_z$ and $V = Y - \hat{Y}_z$ be the respective residuals. Find the correlation coefficient ρ of U and V in terms of the elements of Σ.

c) Based on n i.i.d. observations (X_i, Y_i, Z_i), $i=1,2,\ldots,n$, on (X,Y,Z) find an estimator ρ_n of ρ that will be consistent as $n \to \infty$. State a set of sufficient conditions for this.

d) Establish the asymptotic normality of ρ_n under appropriate conditions.

e) Describe a resampling procedure for approximating the distribution of ρ_n and state sufficient conditions for the validity of such a procedure.

Ph. D. Prelims (Spring 2000) Theory IV (SOLUTION)

a) The best linear est of X given Z is of the form $aZ+b$
Want $\mathrm{Cov}(X-(aZ+b), Z) = 0$ so $a = \frac{\mathrm{Cov}(X,Z)}{VZ}$
Also $EX = aEZ+b$ Since $EX = 0 = EZ$, $b = 0$.
So the best linear est of X given Z $\rho_{X,Z} \frac{\sigma_X}{\sigma_Z}$
where ρ is corr. coeff & σ is std deviation.

b) $U = X - \rho_{XZ} \frac{\sigma_X}{\sigma_Z} Z$, $V = Y - \rho_{YZ} \frac{\sigma_Y}{\sigma_Z} Z$

So $$\mathrm{Cov}(U,V) = \mathrm{Cov}(X,Y) - \cancel{\rho_{XZ}}\, \mathrm{Cov}(X,Z)\, \rho_{YZ} \frac{\sigma_Y}{\sigma_Z} - \mathrm{Cov}(Z,Y)\, \rho_{XZ} \frac{\sigma_X}{\sigma_Z} + \rho_{XZ}\rho_{YZ} \frac{\sigma_X \sigma_Y}{\sigma_Z^2} \sigma_Z^2$$

$$= \mathrm{Cov}(X,Y) - \rho_{XZ}\rho_{YZ}\sigma_X\sigma_Y$$

$V(U) = V(X)(1-\rho_{XZ}^2)$, $V(V) = V(Y)(1-\rho_{YZ}^2)$

So ~~Cov (U~~ $$\rho_{U,V} = \frac{(\rho_{XY} - \rho_{XZ}\rho_{YZ})\sigma_X\sigma_Y}{\sigma_X\sqrt{1-\rho_{XZ}^2}\ \sigma_Y\sqrt{1-\rho_{YZ}^2}}.$$

c) Let F_n be the empirical distribution based on $\{(X_i, Y_i, Z_i) : i = 1, 2, \ldots n\}$.

Denote by $\rho_{U,V}$ in (b) as ρ_F.
Then an estimator of ρ_F is ρ_{F_n}.
Since ρ_{F_n} involve sample Covariances a sufficient condition for strong consistency is finite 2nd moment i.e. $E(X^2+Y^2+Z^2) < \infty$

d) For asymptotic normality (by the delta method) finite 4th moments will do.

e) Resample from F_n, n times to generate (X_i', Y_i', Z_i') $i = 1, 2, \ldots n$ call its empirical cdf F_n'
Compute $\rho_{F_n'}$. Do this N times. Look at the empirical cdf of these $\rho_{F_{n}'}$ values. Finite 4th moment is suff for the validity.