Statistics, the Sample Mean and Variance, and Order Statistics

BA II Year, Madras School of Economics
Instructor: Srikanth B. Pai

Textbook: Larsen & Marx, *An Introduction to Mathematical Statistics and Its Applications*, Fifth Edition.

Topics: statistics as functions of an iid sample; the sample mean and its sampling distribution; the sample variance; order statistics; convergence in probability and almost sure convergence

1. Two chips are drawn without replacement from an urn holding five chips numbered 1 through 5. Let $\bar{X}$ be the average of the two chips drawn. The average of all five chips is 3. Compute $P\big(|\bar{X}-3| > 1.0\big)$.

2. Let $Y_1, Y_2, \ldots, Y_{16}$ be a random sample of size 16 from a normal distribution with mean $\mu = 20$ and $\sigma = 10$. Find the probability that the sample mean $\bar{Y}$ falls between 19.0 and 21.0.

3. Let $X_1, X_2, \ldots, X_n$ be a random sample (iid) with common mean μ and common variance $\sigma^2 < \infty$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Find $E[\bar{X}]$ and $\mathrm{Var}(\bar{X})$.

4. Let $X_1, X_2, \ldots, X_n$ be a random sample (iid) with common mean μ and common variance σ^2, and let the sample variance be
$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$
Show that $E[S^2] = \sigma^2$.

5. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the Uniform$(0, \theta)$ distribution, and let $Y_{(n)}$ be the sample maximum.
   (a) Taking $n = 6$ and $\theta = 3.0$, find $P\big(|Y_{(n)} - \theta| < 0.2\big)$.
   (b) Recompute the probability in part (a) for $n = 3$.

6. Let $2n+1$ observations be drawn as a random sample from the Uniform$(0, 1)$ distribution, and let X be the sample median. Compute $E[X]$.

7. Let $X_1, X_2, \ldots$ be a sequence of random variables.
   (a) Suppose $X_n \xrightarrow{p} 0$. Prove that $\sin X_n \xrightarrow{p} 0$.
   (b) More generally, let g be a continuous function and suppose $X_n \xrightarrow{p} X$. Argue that $g(X_n) \xrightarrow{p} g(X)$.

8. Pick a point ω at random, uniformly, from $[0, 1)$, so that the chance ω lands in a subinterval $[a, b) \subseteq [0, 1)$ equals its length $b - a$. We build a sequence of random variables $X_1, X_2, \ldots$ from ω as follows. For each $k = 0, 1, 2, \ldots$, cut $[0, 1)$ into 2^k equal pieces,
$$I_{k,0} = \left[0, \tfrac{1}{2^k}\right), \ I_{k,1} = \left[\tfrac{1}{2^k}, \tfrac{2}{2^k}\right), \ \ldots, \ I_{k,\,2^k-1} = \left[\tfrac{2^k-1}{2^k}, 1\right).$$

List these intervals one block at a time: first the single piece for $k=0$, then the two pieces for $k=1$, then the four for $k=2$, and so on. Define X_m from the m-th interval on this list by

$$X_m(\omega) = \begin{cases} 1 & \text{if } \omega \text{ lies in the } m\text{-th interval,} \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $X_n \xrightarrow{p} 0$.

(b) Show that X_n does *not* converge to 0 almost surely.

9. Let $U_1, U_2, \ldots$ be a random sample (iid) from the Uniform$(0,1)$ distribution and let $M_n = \max(U_1, \ldots, U_n)$. Prove that $M_n \xrightarrow{p} 1$.

Challenge Problem 1. A single homogeneous good is offered by many sellers, but they do not all charge the same price: a seller's price is an independent draw from a known distribution with cdf F and pdf f. A buyer who wants one unit cannot learn a seller's price without asking, and each price quotation costs him $c > 0$ in time and effort. He decides in advance how many sellers, n, to canvass; he then buys one unit at the lowest of the n prices quoted. Additional search lowers the price he expects to pay but adds to his search cost. Find the amount of search n that minimizes his expected total expenditure, price paid plus cost of search. Evaluate the optimal n when prices are Uniform on $(0,1)$.

Challenge Problem 2. Let $Z_1, Z_2, \ldots$ be a fixed iid sequence, and for each n let $X_{(n)}$ be the sample median of the first $2n+1$ observations $Z_1, \ldots, Z_{2n+1}$. Does $X_{(n)}$ converge in probability? If so, identify its limit.

Answer Key (final answers only)

1. $P\big(|\bar{X} - 3| > 1.0\big) = \dfrac{2}{10} = 0.20.$

2. $\bar{Y} \sim N(20,\ 100/16)$, so $\mathrm{sd}(\bar{Y}) = 2.5$ and $P(19 \le \bar{Y} \le 21) = P(|Z| \le 0.4) \approx 0.311$.

3. $E[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \sigma^2/n.$

4. $E[S^2] = \sigma^2$.

5. (a) $1 - (2.8/3)^6 \approx 0.339$.
   (b) $1 - (2.8/3)^3 \approx 0.187$.

6. $X \sim \mathrm{Beta}(n+1,\ n+1)$, so $E[X] = \dfrac{1}{2}$.

7. (a) $\sin X_n \xrightarrow{p} 0$; use $|\sin x| \le |x|$, so $\{|\sin X_n| > \varepsilon\} \subseteq \{|X_n| > \varepsilon\}$.
   (b) $g(X_n) \xrightarrow{p} g(X)$ (the continuous mapping theorem).

8. (a) For $2^k \le n < 2^{k+1}$, $P(X_n = 1) = 2^{-k}$, which $\to 0$; so $X_n \xrightarrow{p} 0$.
   (b) Every ω lies in exactly one interval per block, so $X_n(\omega) = 1$ for infinitely many n. Thus $X_n(\omega)$ never settles down to 0, for any ω; convergence fails almost surely.

9. $P(|M_n - 1| > \varepsilon) = P(M_n < 1 - \varepsilon) = (1 - \varepsilon)^n \to 0$, so $M_n \xrightarrow{p} 1$.

Challenge Problem. Expected total expenditure with n searches is $E[Y_{(1)}] + cn$, the expected lowest price plus the cost of search; the buyer searches until the expected saving from one more quotation falls to c. For Uniform$(0, 1)$ prices, $E[Y_{(1)}] = \dfrac{1}{n+1}$, and minimizing $\dfrac{1}{n+1} + cn$ gives $n^\star = \dfrac{1}{\sqrt{c}} - 1$ (rounded to a positive integer). For example, $c = 0.01$ gives $n^\star = 9$.

Challenge Problem 2. Yes. $X_{(n)} \xrightarrow{p} m$, where m is the population median (the value with $F(m) = \frac{1}{2}$), provided that median is unique.