Chapter 7: Sums of Random Variables and Long-Term Averages

7.1 Sums of Random Variables

7.1 $\mathcal{E}[X+Y+Z] = \mathcal{E}[X] + \mathcal{E}[Y] + \mathcal{E}[Z] = 0$

a) From Eqn. 5.3 we have

$$\begin{aligned} VAR(X+Y+Z) &= VAR(X) + VAR(Y) + VAR(Z) \\ &\quad + 2COV(X,Y) + 2COV(X,Z) + 2COV(Y,Z) \\ &= 1+1+1+2\left(\frac{1}{4}\right) + 2(0) + 2\left(-\frac{1}{4}\right) = 3 \end{aligned}$$

b) From Eqn. 5.3 we have

$$VAR(X+Y+Z) = VAR(X) + VAR(Y) + VAR(Z) = 3$$

7.2 $\mathcal{E}[S_n] = \mathcal{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathcal{E}[X_i] = n\mu$

$$VAR(S_n) = \underbrace{\sum_{k=1}^{n} VAR(X_k)}_{\substack{\text{sum of diag.} \\ \text{elements of} \\ \text{covariance matrix } K}} + \underbrace{\sum_{j=1}^{n}\sum_{\substack{k=1 \\ j\neq k}}^{n} COV(X_j, X_k)}_{\substack{\text{sum of off-diag.} \\ \text{element of } K}}$$

$$K = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & 0 & & \ldots & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & \ldots & 0 \\ & & & & & 0 \\ & \ddots & \ddots & \ddots & & \rho\sigma^2 \\ & & & \rho\sigma^2 & & \sigma^2 \end{bmatrix}$$

$\therefore VAR(S_n) = n\sigma^2 + 2(n-1)\rho\sigma^2$

7.18 For $n = 16$, Eqn. 7.20 gives $\quad \sigma_X^2 = 1$

$$P[|M_{16} - 0| < \varepsilon] \geq 1 - \frac{1^2}{16\varepsilon^2} = 1 - \frac{1}{16}\frac{1}{\varepsilon^2}$$

Since M_{16} is Gaussian with mean 0 and variance $\frac{1}{16}$

$$\begin{aligned} P[|M_{16} - 0| < \varepsilon] &= P[-\varepsilon < M_{16} < \varepsilon] = 1 - 2Q(\sqrt{16}\varepsilon) \\ &= 1 - 2Q(4\varepsilon) \end{aligned}$$

Similarly for $n = 81$ we obtain

$$\begin{aligned} P[|M_{81} - 0] < \varepsilon] &\geq 1 - \frac{1}{81}\frac{1}{\varepsilon^2} \\ P[|M_{81} - 0] < \varepsilon] &= 1 - 2Q(9\varepsilon) \end{aligned}$$

For example if $\varepsilon = \frac{1}{2}$

$$\begin{aligned} P[|M_{16} < \frac{1}{2}] &\geq 1 - \frac{1}{16/4} = .75 \\ P[|M_{16}| < \frac{1}{2}] &= 1 - 2Q(2) = 1 - 2(2.28 \times 10^{-2}) = 0.954 \\ P[|M_{81}| < \frac{1}{2}] &\geq 1 - \frac{1}{81/4} = .951 \\ P[|M_{81}| < \frac{1}{2}] &= 1 - 2Q(4.5) = 1 - 2(3.4)10^{-6} \end{aligned}$$

Note the significant discrepancies between the bounds and the exact values.

7.19

$$\begin{aligned} P\left[\left|\frac{1}{n}S_n - \mu\right| > \varepsilon\right] &\leq \frac{VAR(\frac{1}{n}S_n)}{\varepsilon^2} = \frac{VAR(S_n)}{n^2\varepsilon^2} \\ &= \frac{n\sigma^2 + 2(n-1)\rho\sigma^2}{n^2\varepsilon^2} \to 0 \quad \text{as } n \to \infty \end{aligned}$$

$\Rightarrow$ Week Law of Large Numbers holds.

7.20

$$\begin{aligned}
P\left[\left|\frac{1}{n}S_n - \mu\right| > \varepsilon\right] &\leq \frac{VAR(S_n)}{n^2\varepsilon^2} \\
&= \frac{1}{n^2\varepsilon^2}\left[n\sigma^2 + 2\rho\sigma^2\left(\frac{n-1}{1-\rho} - \frac{\rho}{1-\rho}\frac{1-\rho^{n-1}}{1-\rho}\right)\right] \\
&= \frac{\sigma^2}{n\varepsilon^2} + \frac{2\rho\sigma^2}{\varepsilon^2}\left(\frac{n-\frac{1}{n}}{n(1-\rho)} - \frac{1}{n^2}\frac{\rho(1-\rho^{n-1})}{(1-\rho)^2}\right) \\
&\to 0 \text{ as } n \to \infty \text{ (assuming } \rho < 1)
\end{aligned}$$

$\Rightarrow$ Weak Law of Large Numbers holds.

7.21 a)

$$\begin{aligned}
LHS &= \sum_{j=1}^{n}(X_j^2 - 2\mu X_j + \mu^2) = \sum_{j=1}^{n} X_j^2 - 2\mu(nM_n) + n\mu^2 \\
RHS &= \sum_{j=1}^{n}(X_j^2 - 2M_nX_j + M_n^2) + n(M_n - \mu)^2 \\
&= \sum_{j=1}^{n} X_j^2 - 2M_n(nM_n) + nM_n^2 + nM_n^2 - 2n\mu M_n + n\mu^2 \\
&= \sum_{j=1}^{n} X_j^2 - 2n\mu M_n + n\mu^2 = LHS \quad \checkmark
\end{aligned}$$

b)

$$\begin{aligned}
\mathcal{E}\left[k\sum_{j=1}^{n}(X_j - M_n)^2\right] &= k\mathcal{E}\Bigg[\underbrace{\sum_{j=1}^{n}(X_j-\mu)^2 - n(M_n-\mu)^2}_{\text{from part a}}\Bigg] \\
&= k\sum_{j=1}^{n}\mathcal{E}[(X_j-\mu)^2] - kn\mathcal{E}[(M_n-\mu)^2] \\
&= kn\sigma^2 - kn\frac{\sigma^2}{n} \\
&= k(n-1)\sigma^2 \quad \text{since } VAR[M_n] = \tfrac{\sigma^2}{n}.
\end{aligned}$$

c) If $k = \frac{1}{n-1}$ then $\mathcal{E}[V_n^2] = \sigma^2$

d) if $k = \frac{1}{n}$ then

$$\mathcal{E}\left[\frac{1}{n}\sum_{j=1}^{n}(X_j - M_n)^2\right] = \left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2 - \underbrace{\frac{1}{n}\sigma^2}_{\text{bias}}$$

7.3 The Central Limit Theorm

7.22 The relevant parameters are $n = 100$, $m = np = 50$, $\sigma^2 = npq = 25$. The Central Limit Theorem then gives:

$$P[40 \le N \le 60] = P\left[\underbrace{\frac{40-50}{\sqrt{25}}}_{-2} \le \frac{N-m}{\sigma} \le \underbrace{\frac{60-50}{\sqrt{25}}}_{2}\right]$$

$$\approx Q(-2) - Q(2) = 1 - 2Q(2) = 1 - 2(2.28)\times 10^{-2} = 0.9544$$

$$P[50 \le N \le 55] \approx Q(0) - Q(1) = \frac{1}{2} - .159 = 0.441$$

7.23 The relative frequency $f_A(n)$ has mean $\frac{2}{10}$ and variance $\frac{1}{n}p(1-p) = \frac{0.16}{n}$

$$P[|f_Z(n) - 0.2| < 0.02] = P[0.18 < f_A(n) < 0.22]$$

$$= P\left[\frac{0.18-0.20}{\sqrt{\frac{0.16}{n}}} < \frac{f_A(n)-0.1}{\sqrt{\frac{0.16}{n}}} < \frac{0.22-0.20}{\sqrt{\frac{0.16}{n}}}\right]$$

$$\approx 1 - 2Q\left(\frac{0.02}{\sqrt{\frac{0.16}{n}}}\right) = 0.95$$

$$\Rightarrow Q\left(\frac{\sqrt{n}}{20}\right) = 0.025 \Rightarrow \frac{\sqrt{n}}{20} = 1.95 \Rightarrow n = 1521$$

7.24

$$S = \sum_{i=1}^{20} X_i \Rightarrow E[S] = 20\, E[X] = 20 \times 3.5 = 70$$

$$VAR[S] = 20\, VAR[X] = 20 \times 2.92 = 58.4$$

Using CLT we have: $S \sim N(70, \sqrt{58.4})$

$$P\{60 < S < 80\} = P\left\{\frac{60-70}{7.64} < \frac{S-70}{7.64} < \frac{80-70}{7.64}\right\}$$

$$= 1 - 2Q(1.3089) = 0.8094$$