M362K: April 3rd, 2024.

Section 3.3. [cont'd].

Problem. Let X and Y be two random variables on the same Ω.
Let
$$\mu_X = \mathbb{E}[X] = 7, \quad \boxed{\sigma_X = 1},$$
$$\mu_Y = \mathbb{E}[Y] = -3, \quad \boxed{\sigma_Y = 2}, \quad \boxed{\rho_{X,Y} = corr[X,Y] = -0.5}.$$
$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \cdot \sigma_Y}$$
Find $Var[2X+Y-5]$.

$\longrightarrow$: $Var[2X+Y\underbrace{-5}_{\text{shift}}] = Var[2X+Y] =$

$$= Var[2X] + 2Cov[2X,Y] + Var[Y]$$
$$= 4\cdot\boxed{Var[X]} + 2\cdot 2\, Cov[X,Y] + \underbrace{Var[Y]}_{=\sigma_Y^2}$$
$$= 4\cdot 1 + 4\cdot(-0.5)\cdot 1\cdot 2 + 2^2$$
$$= 4 - 4 + 4 = 4$$
$\square$

Square Root Law.

Consider $\{X_1, X_2, \dots, X_n, \dots\}$ a sequence of $\boxed{\text{independent, identically distributed}}$ (iid) random variables.

Let $S_n = X_1 + X_2 + \dots + X_n$

$\boxed{\bar{X}_n = \frac{1}{n}\cdot S_n}$ for all $n = 1, 2, \dots$

Focus on S_n:

- $\mathbb{E}[S_n] = \mathbb{E}[X_1 + X_2 + \dots + X_n]$

  linearity of $\mathbb{E}$

  $= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]$

  identically distributed

  $= n\cdot\mu_X$

- $Var[S_n] = Var[X_1 + X_2 + \dots + X_n]$

  independence

  $= Var[X_1] + Var[X_2] + \dots + Var[X_n]$

  identically distributed

  $= n\cdot\sigma_X^2 \implies SD[S_n] = \sigma_X\sqrt{n}$

Focus on $\bar{X}_n$:

- $\mathbb{E}[\bar{X}_n] = \mathbb{E}[\frac{1}{n} \cdot S_n] = \frac{1}{n} \cdot \mathbb{E}[S_n] = \frac{1}{n} \cdot n \cdot \mu_X = \mu_X$
- $\mathrm{Var}[\boxed{\bar{X}_n}] = \mathrm{Var}[\frac{1}{n} S_n] = \frac{1}{n^2} \mathrm{Var}[S_n]$
  $= \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \sigma_X^2 = \frac{\sigma_X^2}{n}$
- $SD[\boxed{\bar{X}_n}] = \frac{\sigma_X}{\sqrt{n}}$

Law of Averages (aka The Law of Large Numbers (LLN))

$$\boxed{\frac{X_1 + X_2 + \cdots + X_n}{n}} \xrightarrow[n \to \infty]{} \mu_X$$

The Normal Approximation Theorem. (aka The Central Limit Theorem (CLT))

$$\boxed{\frac{S_n - n \cdot \mu_X}{\sigma_X \cdot \sqrt{n}} \xrightarrow[n \to \infty]{\mathcal{D}} N(0,1)}$$

Then,

$$\mathbb{P}[c \leq S_n \leq d] = \mathbb{P}\left[\frac{c - n \cdot \mu_X}{\sigma_X \sqrt{n}} \leq \underbrace{\boxed{\frac{S_n - n \cdot \mu_X}{\sigma_X \sqrt{n}}}}_{\approx N(0,1) \sim Z} \leq \frac{d - n \cdot \mu_X}{\sigma_X \sqrt{n}}\right]$$

$$\approx \mathbb{P}\left[\frac{c - n \cdot \mu_X}{\sigma_X \sqrt{n}} \leq Z \leq \frac{d - n \cdot \mu_X}{\sigma_X \sqrt{n}}\right]$$

$$= \Phi\left(\frac{d - n \cdot \mu_X}{\sigma_X \sqrt{n}}\right) - \Phi\left(\frac{c - n \cdot \mu_X}{\sigma_X \sqrt{n}}\right)$$

Remark. We must use the continuity correction for r.v.s whose support contains consecutive integers.

Problem 3.3.24. from Pitman.

A box contains 4 tickets numbered 0, 1, 1, 2.
Let S_n be the sum of the numbers obtained from n draws w/ replacement @ random.

A single draw X has the pmf

x	0	1	2
$p_X(x)$	1/4	1/2	1/4

The sum S_2 of two independent draws has the pmf

s	0	1	2	3	4
$p_{S_2}(s)$	1/16	1/4	3/8	1/4	1/16

$S_2 = 1$ iff $X_1 = 0$ and $X_2 = 1$
OR
$X_1 = 1$ and $X_2 = 0$

$$\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{4} = 1/4$$

$S_2 = 3$ same by symmetry

Using the CLT, approximate $\boxed{\mathbb{P}[S_{50} = 50]}$

We need:
- $\mu_X = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} = 1$
- $\sigma_X = \sqrt{Var[X]}$

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$\mathbb{E}[X^2] = 0^2\cdot\frac{1}{4} + 1^2\cdot\frac{1}{2} + 2^2\cdot\frac{1}{4} = \frac{3}{2}$$

$$\sigma_X = \sqrt{\frac{3}{2} - 1} = \frac{\sqrt{2}}{2}$$

We do need the continuity correction!

$$\Phi\left(\frac{50.5-50}{\underbrace{\frac{\sqrt{2}}{2}\cdot\sqrt{50}}_{5}}\right) - \Phi\left(\frac{49.5-50}{\underbrace{\frac{\sqrt{2}}{2}\cdot\sqrt{50}}_{5}}\right) = \Phi(0.10) - \overbrace{\Phi(-0.10)}^{\substack{\text{Symmetry of } N(0,1) \\ 1-\Phi(0.1)}}$$

$$= 2\cdot\Phi(0.10) - 1$$

$$= 2\cdot 0.5398 - 1 = 0.0796 \quad \square$$