Variance, Covariance and Central Limit Theorem.

1. *A doctor prepared bills for 10 patients. Her absent minded secretary put 10 bills into 10 addressed envelops randomly. Let $X_i = 1$ if the i-th bill was sent ot a correct address.*

(a) Compute EX_1, VX_1, $Cov(X_1, X_2)$,

(b) Let X be number of bills sent to the correct address. Compute $E(X)$ and $V(X)$.

Solution. The distribution of X_1 is the following

X	0	1
P	$\frac{9}{10}$	$\frac{1}{10}$

Thus $EX_1 = \frac{1}{10}$ and $V(X_1) = \frac{1}{10} \times \frac{9}{10} = \frac{9}{100}$.

Next

$$P(X_1X_2 = 1) = P(X_1 = 1 \text{ and } X_2 = 1) = \frac{1}{10 * 9} = \frac{1}{90}$$

since there are 10 ways to select the bill for the first patient and 9 ways to select the bill for the second patient. Thus the distribution of X_1X_2 is the following

X_1X_2	0	1
P	$\frac{89}{90}$	$\frac{1}{90}$

and so $E(X_1X_2) = \frac{1}{90}$ and

$$\text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \frac{1}{90} - \frac{1}{100} = \frac{1}{900}.$$

Next

$$E(X) = \sum_{j=1}^{10} E(X_j) = 10 \times \frac{1}{10} = 1,$$

$$V(X) = \sum_{j=1}^{10} V(X_j) + 2\sum_{i<j} \text{Cov}(X_iX_j) = 10 \times \frac{9}{100} + 2\left(\begin{array}{c}10\\2\end{array}\right)\frac{1}{900} = \frac{9}{10} + \frac{90}{900} = \frac{9}{10} + \frac{1}{10} = 1.$$

2. *Compute EX and VX if*

(a) $X \sim Bin(n, p)$, (b) $X \sim Hyp(N, M, n)$, (c) $X \sim NBin(r, p)$.

Solution. (a) $X = \sum_{j=1}^n X_j$ where $X_j = 1$ if trial j is success and 0 otherwise. The distribution of X_j is

X_j	0	1
P	1-p	p

Thus $E(X_j) = 0 \times (1-p) + 1 \times p = p$. Likewise $E(X_j^2) = p$ and so $V(X_j) = p - p^2 = p(1-p)$. Since X_j are independent

$$EX = \sum_{j=1}^n E(X_j) = np, \quad VX = \sum_{j=1}^n V(X_j) = np(1-p).$$

(b) $X = \sum_{j=1}^n X_j$ where $X_j = 1$ if ball j is white and 0 otherwise. The distribution of X_j is

X_j	0	1
P	$\frac{N-M}{N}$	$\frac{M}{N}$

Thus $E(X_j) = 0 \times \frac{N-M}{M} + 1 \times \frac{M}{N} = \frac{M}{N}$. Likewise $E(X_j^2) = \frac{M}{N}$ and so

$$V(X_j) = \frac{M}{N} - \left(\frac{M}{N}\right)^2 = \frac{M(N-M)}{N^2}.$$

Next

$$P(X_iX_j = 1) = P(X_i = 1 \text{ and } X_j = 1) = P(W_iW_j) = P(W_i)P(W_j|W_i) = \frac{M}{N}\frac{M-1}{N-1}.$$

Accordingly

$$E(X_iX_j) = 1 \times P(X_iX_j = 1) + 0 \times P(X_iX_j = 0) = \frac{M}{N}\frac{M-1}{N-1}.$$

Accordingly

$$Cov(X_i, X_j) = \frac{M}{N}\frac{M-1}{N-1} - \left(\frac{M}{N}\right)^2 = \frac{M}{N}\left(\frac{M}{N} - \frac{M-1}{N-1}\right) = -\frac{M(N-M)}{N^2(N-1)}.$$

Therefore

$$EX = \sum_{j=1}^{n} E(X_j) = \frac{Mn}{N},$$

$$VX = \sum_{j=1}^{n} V(X_j) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

$$= \frac{nM(N-M)}{N^2} - 2\binom{n}{2}\frac{M(N-M)}{N^2(N-1)} = \frac{nM(N-M)}{N^2}\left(1 - \frac{n-1}{N-1}\right) = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

(c) Denote $q = 1-p$. Let X_k be the number of failures between (k-1)-st and k-th successes. Then X_k are independent and have geometric distribution with parameter p. Thus $P(X_k = n) = q^n p$. Hence

$$E(X_k) = \sum_{n=0}^{\infty} nq^n p = pq\sum_{n=0}^{\infty} nq^{n-1} = pq\left(\sum_{n=0}^{\infty} q^n\right)' = pq\left(\frac{1}{1-q}\right)' = \frac{pq}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}.$$

$$E(X_k^2) = \sum_{n=0}^{\infty} n^2 q^n p = \sum_{n=0}^{\infty} nq^n p + \sum_{n=0}^{\infty} n(n-1)q^n p = \frac{p}{q} + pq^2\sum_{n=0}^{\infty} n(n-1)q^{n-2} = \frac{p}{q} + pq^2\left(\sum_{n=0}^{\infty} q^n\right)''$$

$$= \frac{p}{q} + pq^2\left(\frac{1}{1-q}\right)'' = \frac{p}{q} + \frac{2pq^2}{(1-q)^3} = \frac{p}{q} + \frac{2pq^2}{p^3} = \frac{q}{p} + \frac{2q^2}{p^2}.$$

Thus

$$V(X_k) = \frac{q}{p} + \frac{2q^2}{p^2} - \frac{q^2}{p^2} = \frac{q}{p} + \frac{q^2}{p^2} = \frac{pq+q^2}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2}.$$

Now $X = \sum_{k=1}^{r} X_k$. Hence

$$EX = \sum_{k=1}^{r} E(X_k) = \frac{rq}{p}, \quad V(X) = \sum_{k=1}^{r} V(X_k) = \frac{rq}{p^2}.$$

3. *Suppose 50 % of all drivers have American cars, 40 % have Japanese cars and 10 % have European cars. Consider 15 consecutive cars crossing certain intersection. Let A be the number of American cars among those 15 cars and J be the number of Japanese cars. Compute $Cov(A, J)$, $Corr(A, J)$.*

Solution. Let $A_j = 1$ if car j is American and $A_j = 0$ otherwise, $J_j = 1$ if car j is Japanese and $J_j = 0$ otherwise. Then $A_j \sim$Bern(0.5), $J_j \sim$Bern(0.4). Thus $E(A_j) = 0.5$ and $E(J_j) = 0.4$ Also $A_j J_j = 0$ since a car can not be American and Japanese at the same time. Hence

$$\mathrm{Cov}(A_j, J_j) = E(A_j J_j) - E(A_j)A(J_j) = 0 - 0.5 \times 0.4 = -0.2.$$

On the other hand for $i \neq j$ we have $\mathrm{Cov}(A_i, J_j) = 0$ by independence. Hence

$$\mathrm{Cov}(A, J) = \mathrm{Cov}\left(\sum_{i=1}^{15} A_i, \sum_{j=1}^{15} J_j\right) = \sum_{i\neq j} \mathrm{Cov}(A_i, J_j) + \sum_{j=1}^{15} \mathrm{Cov}(A_j, J_j) = -0.2 \times 15 = -3.$$

Therefore

$$Corr(A, J) = \frac{Cov(A, J)}{\sqrt{V(A)V(J)}} = \frac{-3}{\sqrt{15 \times (0.5)^2 \times 15 \times 0.4 \times 0.6}} \approx -0.82.$$

4. *Fruitland's economy consists of two companies. Each issue shares costing 10 oranges/share. During the next year Maracuja's shares will grow 2 oranges with probability 1/2 or drop 1 orange with probability 1/2. PassionFriuit's shares will grow 3 oranges with probability 1/2 or drop 2 oranges with probability 1/2 independent of Maracuja. You have 340 oranges (an orange is the name of Fruitland's currency so you can not eat it). How much should you invest into each company to minimize you risk?*

Solution. The gain on Maracuja shares has the following distribution.

G_1	-1	2
P	$\frac{1}{2}$	$\frac{1}{2}$

Thus

$$E(G_1) = -1 \times \frac{1}{2} + 2 \times \frac{1}{2} = \frac{1}{2}, \quad E(G_1^2) = 1 \times \frac{1}{2} + 4 \times \frac{1}{2} = \frac{5}{2}, \quad V(G_1) = \frac{5}{2} - \frac{1}{4} = \frac{9}{4}.$$

The gain on Passion Fruit shares has the following distribution.

G_2	-2	3
P	$\frac{1}{2}$	$\frac{1}{2}$

Thus

$$E(G_2) = -2 \times \frac{1}{2} + 3 \times \frac{1}{2} = \frac{1}{2}, \quad E(G_1^2) = 4 \times \frac{1}{2} + 9 \times \frac{1}{2} = \frac{13}{2}, \quad V(G_1) = \frac{13}{2} - \frac{1}{4} = \frac{25}{4}.$$

Suppose you buy x shares of Maracuja and $34 - x$ of Passion Fruit. Then your gain is $G = xG_1 + (34 - x)G_2$. Hence

$$V(G) = x^2 V(G_1) + (34 - x)^2 V(G_2) = \frac{9}{4}x^2 + \frac{25}{4}(34 - x)^2.$$

Considered as a function of x this is a parabala with the vertex at x where

$$\frac{9}{2}x - \frac{25}{2}(34 - x) = 0, \text{ that is } 25 \times 34 = (25 + 9)x \text{ or } x = \frac{25 \times 34}{34} = 25.$$

Hence one needs to by 25 shares of Maracuja and 9 shares of Passion Fruit.

5. *John takes bus to work 300 days a year. Suppose that bus waiting times are independent of one another and each is uniformly distributed between 0 and 10 min. Let T be total waiting time for a particular year.*

(a) Compute $E(T)$ and $V(T)$.

(b) Compute approximately the probability that T is more than 26 hours; that it is between 24 and 25 hours.

Solution. Let T_j be the waiting time during day j. Then $T = \sum_{j=1}^{300} T_j$. Next $ET_j = \frac{10}{2} = 5$ and $V(T_j) = \frac{10^2}{12} = \frac{25}{3}$. Hence

$$E(T) = 300 \times 5 = 1500 \text{ and } V(T) = 300 \times \frac{25}{3} = 2500 = 50^2.$$

By the Central Limit Theorem $T \approx \mathcal{N}(1500, 50^2)$ and so

$$P(T > 1560) \approx 1 - \Phi\left(\frac{1560 - 1500}{50}\right) = 1 - \Phi(1.2) \approx 1 - 0.88 = 0.12;$$

$$P(1440 < T < 1500) = P(T < 1500) - P(-1440) \approx \Phi\left(\frac{1500 - 1500}{50}\right) - \Phi\left(\frac{1440 - 1500}{50}\right)$$
$$= 0.5 - \Phi(-1.2) \approx 0.38.$$

6. *A certain lamp uses lightbulbs whose lifetime have exponential distribution with mean 100 hours. A storage room has a supply of 100 bulbs. Find the probability that this supply would not suffice for 9500 hours of work.*

Solution. Let T_j be the lifetime of bulb j. Then $T = \sum_{j=1}^{100} T_j$ is the total lifetime of bulbs in the storage. Then T_j ~EXP(λ) where $E(T_j) = \frac{1}{\lambda} = 100$. Thus $\lambda = 0.01$ and so $V(T_j) = \frac{1}{\lambda^2} = 10000$. Accrdingly

$$ET = 100 \times 100 = 10000 \text{ and } V(T) = 100 \times 10000 = 1000000 = (1000)^2.$$

By the Central Limit Theorem $T \approx \mathcal{N}(10000, (1000)^2)$ and so

$$P(T < 9500) \approx \Phi\left(\frac{9500 - 10000}{1000}\right) = \Phi(-0.5) \approx 0.31.$$

7. *Let X ~Pois(100). Compute approximately $P(100 \leq X \leq 105)$.*

Solution. $X = \sum_{j=1}^{n} X_j$ where X_j ~Pois(1). Thus by the Central Limit Theorem $X \approx \mathcal{N}(100, 100)$ and using continuity correction we get

$$P(100 \leq X \leq 105) = P(X \leq 105) - P(X \leq 99) \approx \Phi\left(\frac{105.5 - 100}{10}\right) - \Phi\left(\frac{99.5 - 100}{10}\right)$$
$$= \Phi(0.55) - \Phi(-0.05) \approx 0.71 - 0.48 = 0.23.$$

Using Poisson distribution calculator we find that

$$P(100 \leq X \leq 105) = P(X \leq 105) - P(X \leq 99) \approx 0.7129 - 0.4867 = 0.2262.$$

8. *Let (X_1, X_2) be a normal vector with mean $(1, 2)$ and covariance matrix $\begin{pmatrix} 1 & 2 \\ 2 & 7 \end{pmatrix}$. Compute*

(a) $P(X_2 > X_1)$.
(b) $P(X_2 > 2 | X_1 = 1)$.

Solution. Let $Y = X_2 - X_1$. Then

$$EY = 2 - 1 = 1 \text{ and } V(Y) = V(X_2) + V(X_1) - 2\text{Cov}(X_2, X_1) = 7 + 1 - 4 = 4.$$

Thus $Y \sim \mathcal{N}(1, 2^2)$. Hence

$$P(Y > 0) = 1 - \Phi\left(\frac{0 - 1}{2}\right) = 1 - \Phi(-0.5) \approx 1 - 0.31 = 0.69.$$

Write $Z = X_2 - cX_1$. Then $\text{Cov}(Z, X_1) = \text{Cov}(X_2, X_1) - cV(X_1) = 2 - c$. Take $c = 2$. Then $\text{Cov}(Z, X_1) = 0$ and so Z and X_1 are independent. Next since $Z = X_2 - 2X_1$ it follows that $EZ = EX_2 - 2EX_1 = 0$ and $VZ = V(X_2) + 4V(X_1) - 4\text{Cov}(X_1, X_2) = 7 + 4 - 8 = 3$. Hence $Z \sim \mathcal{N}(0, 3)$. Therefore $P(X_2 > 2 | X_1 = 1) = P(Z > 0 | X_1 = 1) = P(Z > 0) = \frac{1}{2}$ where the second equality is because Z and X_1 are independent.