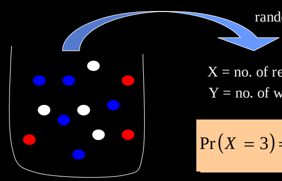
Joint Distribution



randomly draw 3 balls

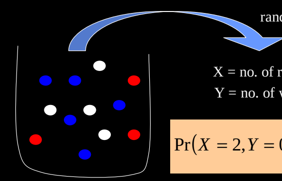
X = no. of red balls {0,1,2,3}
Y = no. of white balls {0,1,2,3}

$$\Pr(X = 3) = \frac{_3C_3}{_{12}C_3} = 0.004545$$

X	0	1	2	3
Prob.	0.3818	0.4909	0.1227	0.004545

Y	0	1	2	3
Prob.	0.2545	0.5091	0.2182	0.01818

Joint Distribution



randomly draw 3 balls

X = no. of red balls {0,1,2,3}
Y = no. of white balls {0,1,2,3}

$$\Pr(X = 2, Y = 0) = \frac{_3C_2 \times _5C_1}{_{12}C_3} = 0.06818$$

X \ Y	0	1	2	3
0	0.04545	0.1818	0.1364	0.01818
1	0.1364	0.2727	0.08182	0
2	0.06818	0.05455	0	0
3	0.004545	0	0	0

Joint Probability Mass Function

$$p(x, y) = \Pr(X = x, Y = y)$$ —— Joint pmf

Example : $p(x,y) = \frac{_3C_x \times _4C_y \times _5C_{3-x-y}}{_{12}C_3} \quad , 0 \le x + y \le 3$

X \ Y	0	1	2	3	Total	
0	0.04505	0.1818	0.1364	0.01818	0.3818	← $\Pr(X = 0)$
1	0.1364	0.2727	0.08182	0	0.4909	← $\Pr(X = 1)$
2	0.06818	0.05455	0	0	0.1227	← $\Pr(X = 2)$
3	0.04545	0	0	0	0.04545	← $\Pr(X = 3)$
Total	0.2545	0.5091	0.2182	0.01818	1	
	↑ $\Pr(Y = 0)$	↑ $\Pr(Y = 1)$	↑ $\Pr(Y = 2)$	↑ $\Pr(Y = 3)$		

Marginal Probability Mass Function

$$p_X(x) = \Pr(X = x) = \sum_y p(x, y)$$

$$p_Y(y) = \Pr(Y = y) = \sum_x p(x, y)$$

Example :

$$p_X(x) = \frac{_3C_x \times _9C_{3-x}}{_{12}C_3} \quad , x = 0,1,2,3$$

$$p_Y(y) = \frac{_4C_y \times _9C_{3-y}}{_{12}C_3} \quad , y = 0,1,2,3$$

Independence

X and Y are independent if

$$p(x, y) = p_X(x) p_Y(y)$$ **for all x,y**

Example :

X \ Y	0	1	2	3	Total
0	0.1190	0.4762	0.3571	0.0476	1
1	0.2778	0.5556	0.1667	0	1
2	0.5556	0.4444	0	0	1
3	1	0	0	0	1

$$p(y \mid x) = \frac{p(x, y)}{p_X(x)}$$ —— Conditional pmf

Independence

Example :

X \ Y	10	20	40	80	$p_X(x)$
20	0.04	0.08	0.08	0.05	0.25
40	0.12	0.24	0.24	0.15	0.75
$p_Y(y)$	0.16	0.32	0.32	0.20	1

$$0.20 \times 0.75 = 0.15$$

X and Y are independent !

Mathematical Expectation

Example : (X, Y) = height and weight of a random man

Joint distribution

X \ Y	75	80	85	90	$p_X(x)$
1.7	0.1	0.08	0.06	0.03	0.27
1.8	0.09	0.2	0.15	0.05	0.49
1.9	0.02	0.05	0.07	0.1	0.24
$p_Y(y)$	0.21	0.33	0.28	0.18	1

$E(X) = 1.797$

$E(Y) = 82.15$

$$E(36X + 0.8Y) = (36\times1.7 + 0.8\times75)(0.1) + \cdots (36\times1.9 + 0.8\times90)(0.1)$$

$$\left(X^2\right) \quad \left(1.7^2 = 130.412 \quad \left(1.9^2\right) \quad \left(1.9^2\right) \quad 5$$

$$= 25.51 = 36E(X) + 0.8E(Y)$$

$$\neq E(Y)/E(X)^2$$

Mathematical Expectation

$E(3X + 4Y) = 3E(X) + 4E(Y)$

$E(7\log X - 2Y^2 + 5\sqrt{XY}) = 7E(\log X) + -2E(Y^2) + 5E(\sqrt{XY})$

In general, $E(XY) \neq E(X)E(Y)$ $\quad E(X/Y) \neq E(X)/E(Y)$

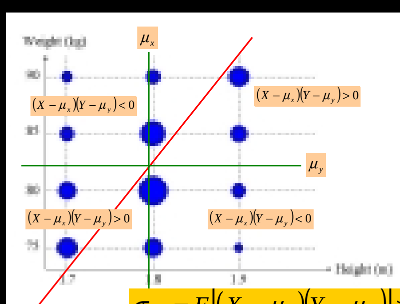
If X and Y are independent, then

$$E(XY) = E(X)E(Y)$$

$$E(X^2\sqrt{Y}) = E(X^2)E(\sqrt{Y})$$

$$E(X/Y) = E(X)E(1/Y)$$

Covariance



m man

$\mu_x = 1.797$

$\mu_y = 82.15$

$$\sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)] > 0$$

Covariance

$$\sigma_{xy} = Cov(X,Y) = E(XY) - \mu_x\mu_y$$

Example : (X, Y) = height and weight of a random man

Joint distribution

X \ Y	75	80	85	90	$p_X(x)$
1.7	0.1	0.08	0.06	0.03	0.27
1.8	0.09	0.2	0.15	0.05	0.49
1.9	0.02	0.05	0.07	0.1	0.24
$p_Y(y)$	0.21	0.33	0.28	0.18	1

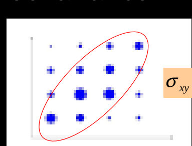
$E(X) = 1.797$

$E(Y) = 82.15$

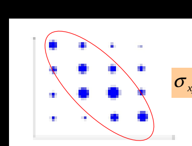
$E(XY) = (1.7)(75)(0.1) + (1.7)(80)(0.08) + \cdots + (1.9)(90)(0.1) = 147.745$

$Cov(X,Y) = 147.745 - (1.797)(82.15) = 0.12145$ +ve correlated

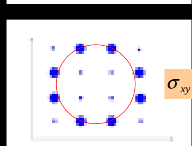
Covariance



$\sigma_{xy} > 0$

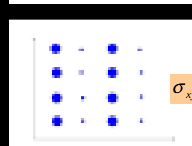


$\sigma_{xy} < 0$



$\sigma_{xy} = 0$

Dependent



$\sigma_{xy} = 0$

Independent

Covariance

X and Y independent $\Rightarrow$ $Cov(X,Y) = 0)E(Y)$

$\Leftarrow$?

Example :

Income ($1000)

No. of dates	10	20	30	Marginal
0	1/3	0	1/3	2/3
1	0	0	0	0
2	0	1/3	0	1/3
Marginal	1/3	1/3	1/3	1

$E(X) = \frac{2}{3}\times0 + \frac{1}{3}\times2 = \frac{2}{3}$

$E(Y) = \frac{1}{3}(10 + 20 + 30) = 20$

$E(XY) = \frac{1}{3}(0\times10 + 0\times30 + 2\times20) = \frac{40}{3}$

X and Y are uncorrelated

X and Y are not independent!

Correlation Coefficient

Magnitude of σ_{xy} depends on the scale of X and Y.

$$Cov(aX + b, cY + d) = acCov(X, Y)$$

Standardized by standard deviations of X and Y :

$$\rho = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$ — Correlation coefficient

$$-1 \le \rho \le 1$$

$$Corr(aX + b, cY + d) = sign(ac)Corr(X, Y)$$

X and Y independent ~~⇔~~ $\rho = 0$

Correlation Coefficient

Example : (X, Y) = height and weight of a random man

Joint distribution

X \ Y	75	80	85	90	$p_X(x)$
1.7	0.1	0.08	0.06	0.03	0.27
1.8	0.09	0.2	0.15	0.05	0.49
1.9	0.02	0.05	0.07	0.1	0.24
$p_Y(y)$	0.21	0.33	0.28	0.18	1

$E(X) = 1.797$

$E(Y) = 82.15$

$\sigma_{xy} = 0.12145$

$$Var(X) = 3.2343 - 1.797^2 = 0.005091 .9^2)(0.24) = 3.2343$$

$$Var(Y) = 6774.25 - 82.15^2 = 25.6275^2)(0.28) + (90^2)(0.18) = 6774.25$$

$$\rho = \frac{0.12145}{\sqrt{(0.005091)(25.6275)}} = 0.3362$$

Slightly +ve correlated

Linear Combination of Random Variables

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$$

Example : X = husband's income Y = wife's income

$E(X) = 20$ $E(Y) = 16$

$Var(X) = 60$ $Var(Y) = 70$ $Cov(X, Y) = 49$

Total income : S = X + Y

$$E(S) = E(X) + E(Y) = 20 + 16 = 36$$

$$Var(S) = 60 + 70 + 2 \times 49 = 228 ;ov(X, Y)$$ $\sigma_S = \sqrt{228} = 15.1$

Linear Combination of Random Variables

Example: investment

B \ S	-10%	0%	10%	20%	Margin
6%	0	0	0.1	0.1	0.2
8%	0	0.1	0.3	0.2	0.6
10%	0.1	0.1	0	0	0.2
Margin	0.1	0.2	0.4	0.3	1

$E(S) = 9\%$

$Var(S) = 89\%^2$

$E(B) = 8\%$

$Var(B) = 1.6\%^2$

$Cov(S, B) = -8\%^2$

$$R = pS + (1-p)B \times 8 = -8)(0) + \cdots + (20)(10)(0) = 64$$ Portfolio

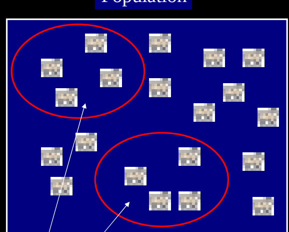
$$E(R) = 8 + p3(1-p)p)E(B)$$

$$Var(R) = 106.6p^2 - 19.2p + 1.6;p(1-p))(1-p)Cov(S, B)$$

$p = 0.09$ $E(R) = 8.09$ $\sigma_R = 0.8576$

Population and Sample

Population



μ, σ^2 — Population parameters

$\bar{X}, S^2$ —Inference→ μ, σ^2

Sample

$\bar{X}, S^2$ — Sample Statistics

Population and Sample

Population

$\mu = 6$

$\sigma^2 = 8$

Sample (with replacement)	Sample Mean	Sample Variance	
$\{X_1, X_2\}$	$\bar{X} = \frac{X_1 + X_2}{2}$	$S^2 = \frac{(X_1 - X_2)^2}{2}$	
2, 2	2	0	Prob = 1/25
2, 4	3	2	Prob = 2/25
4, 8	6	8	Prob = 2/25

……………….

Population and Sample

Sampling distributions

$\overline{X}$	2	3	4	5	6	7	8	9	10
Prob	0.04	0.08	0.12	0.16	0.2	0.16	0.12	0.08	0.04

S^2	0	2	8	18	32
Prob	0.2	0.32	0.24	0.16	0.08

$E(\overline{X}) = 6 = \mu$)4 $+ 3\times 0.08 + \cdots + 10\times 0.04$

Unbiased

$E(S^2) = 8 = \sigma^2$ $+ 2\times 0.32 + \cdots + 32\times 0.08$

$Var(\overline{X}) = 4 \Rightarrow \sqrt{Var(\overline{X})} = 2$

$Var(S^2) = 86.4 \Rightarrow \sqrt{Var(S^2)} = 9.295$

Very Simple Random Sample (VSRS)

$\{X_1, X_2, \ldots, X_n\}$ — (very simple) random sample

- Each X drawn from same population (distribution)
- X’s are independent

$E(\overline{X}) = \mu$

$Var(\overline{X}) = \frac{\sigma^2}{n}$

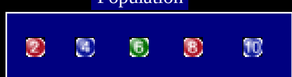
$SE = \sqrt{Var(\overline{X})} = \frac{\sigma}{\sqrt{n}}$ — standard error

Example : Final examination scores $\mu = 71.8$ $\sigma^2 = 195.2$

For a VSRS of 16 students : $E(\overline{X}) = \mu = 71.8$ $SE = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{195.2}{16}} = 3.5$

Sampling Distribution

Population



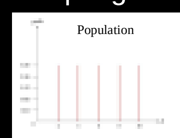
VSRS → $\{X_1, X_2\}$

Sampling distributions

$\overline{X}$	2	3	4	5	6	7	8	9	10
Prob	0.04	0.08	0.12	0.16	0.2	0.16	0.12	0.08	0.04

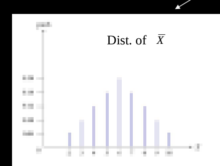
S^2	0	2	8	18	32
Prob	0.2	0.32	0.24	0.16	0.08

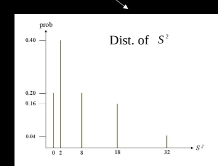
Sampling Distribution



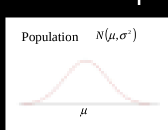
VSRS → $\{X_1, X_2\}$

$\overline{X}$ S^2



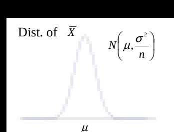


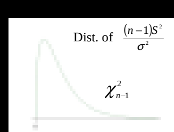
Normal Population



VSRS → $\{X_1, X_2, \ldots, X_n\}$

$\overline{X}$ S^2





Normal Population

Example : measurement X , true value μ
$X \sim N(\mu, 0.01)$

$\Pr(|X - \mu| > 0.05) = \Pr\left(\left|\frac{X-\mu}{0.1}\right| > \frac{0.05}{0.1}\right)$ $= 2(1 - \Phi(0.5)) = 0.617$

Take 10 measurements independently. (VSRS with n = 10)

$\overline{X} \sim N\left(\mu, \frac{0.01}{10}\right)$

$\Pr(|\overline{X} - \mu| > 0.05) = \Pr\left(\left|\frac{\overline{X}-\mu}{0.1/\sqrt{10}}\right| > \frac{0.05}{0.1/\sqrt{10}}\right)$ $= 2(1 - \Phi(1.58)) = 0.1142$

Simple Random Sample (SRS)

$\{X_1, X_2, ..., X_n\}$ — simple random sample

- Sampling without replacement from population (size N)
- Equal probability for each possible sample

$E(\bar{X}) = \mu$ $\quad Var(\bar{X}) = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n}$ $\quad SE = \sqrt{Var(\bar{X})} = \sqrt{\frac{N-n}{N-1}}\frac{\sigma}{\sqrt{n}}$ — standard error

$\frac{N-n}{N-1} \le 1$ — Finite population correction factor

Simple Random Sample

Population

$\mu = 6$

$\sigma^2 = 8$

Sample (**without** replacement)	Sample Mean	Sample Variance	
$\{X_1, X_2\}$	$\bar{X} = \frac{X_1 + X_2}{2}$	$S^2 = \frac{(X_1 - X_2)^2}{2}$	
2, 4	3	2	Prob = 1/10
2, 6	4	8	Prob = 1/10
4, 8	6	8	Prob = 1/10

………………

Simple Random Sample

Sampling distributions

$\bar{X}$	3	4	5	6	7	8	9
Prob	0.1	0.1	0.2	0.2	0.2	0.1	0.1

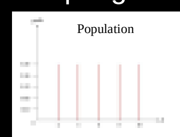
S^2	2	8	18	32
Prob	0.4	0.3	0.2	0.1

$E(\bar{X}) = 6 = \mu\ 0.1 + 4\times($ Unbiased 0.1

$E(S^2) = 10 \ne \sigma^2 + 8\times 0.$ Biased 2×0.1

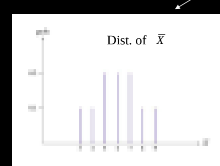
$Var(\bar{X}) = 3 \Rightarrow \sqrt{Var(\bar{X})} = 1.732$ $\quad Var(S^2) = 88 \Rightarrow \sqrt{Var(S^2)} = 9.38$

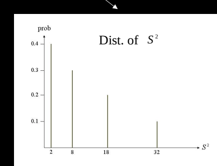
Sampling Distribution (SRS)



SRS → $\{X_1, X_2\}$

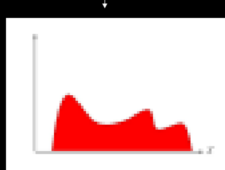
$\bar{X}$ $\quad S^2$

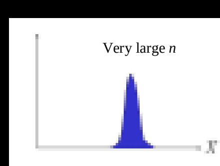




Central Limit Theorem

Arbitrary population
(mean μ, variance σ^2) — VSRS → $\{X_1, X_2, ..., X_n\}$ → $\bar{X}$





Central Limit Theorem

Arbitrary population
(mean μ, variance σ^2) — VSRS → $\{X_1, X_2, ..., X_n\}$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{L} N(0,1) \quad \text{as } n \to \infty$$

For large n $\quad \Pr(\bar{X} \le c) \approx \Phi\left(\frac{c-\mu}{\sigma/\sqrt{n}}\right)$ — Normal approximation

~~$\Pr(X \le c) \overset{?}{\approx} \Phi\left(\frac{c-\mu}{\sigma}\right)$~~

Normal Approximation

Example : Monthly income

Income	4000	7000	10000	15000	20000	30000	40000	60000
Prob	0.4	0.25	0.15	0.1	0.05	0.03	0.01	0.01

$\mu = E(X) = 9250$ $(0.4)+(7000)(0.2$ $\sigma = \sqrt{Var(X)} = 8342$

For a VSRS with size 100 $\bar{X} \sim N\left(9250, \frac{8342^2}{100}\right)$

$\Pr(\bar{X} > 10000) \approx 1 - \Phi\left(\frac{10000-9250}{8342/10}\right) = 1 - \Phi(0.90) = 0.1841$

$\Pr(X > 10000) = 0.1 + 0.05 + 0.03 + 0.01 + 0.01 = 0.2$

Normal Approximation

Normal approximate binomial

$Y \sim b(n,\pi) \xrightarrow{n \text{ large}} \frac{Y - n\pi}{\sqrt{n\pi(1-\pi)}} \sim N(0,1)$

Example : $Y \sim b(35, 0.25)$

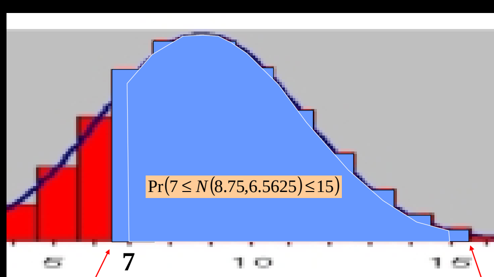
$E(Y) = 35 \times 0.25 = 8.75$ $Var(Y) = 35 \times 0.25 \times 0.75 = 6.5625$

$\Pr(7 \le Y \le 15) \approx 0.9927 - (1 - 0.7527) = 0.7454$

By exact calculation,

$\Pr(7 \le Y \le 15) = \sum_{y=7}^{15} {}_{35}C_y (0.25)^y (0.75)^{35-y} = 0.8018$

Normal Approximation



Continuity Correction

Approximate discrete distribution by continuous distribution

$\Pr(X \le c) \rightarrow \Pr(X \le c + 0.5)$

$\Pr(X < c) \rightarrow \Pr(X \le c - 0.5)$

$\Pr(X \ge c) \rightarrow \Pr(X \ge c - 0.5)$

$\Pr(X > c) \rightarrow \Pr(X \ge c + 0.5)$

Example : $Y \sim b(35, 0.25)$ $E(X) = 8.75$ $Var(X) = 6.5625$

$\Pr(7 \le Y \le 15) \approx 0.9958 - (1 - 0.8100) = 0.8058$

By exact calculation, $\Pr(7 \le Y \le 15) = 0.8018$

Normal Approximation

Normal approximate Poisson

$Y \sim \wp(\theta) \xrightarrow{\theta \text{ large}} \frac{Y - \theta}{\sqrt{\theta}} \sim N(0,1)$

Example : $Y \sim \wp(30)$

By normal approximation with continuity correction,

$\Pr(24 \le Y \le 39) \approx 0.9586 - (1 - 0.8824) = 0.841$

By exact calculation,

$\Pr(24 \le Y \le 39) = \sum_{y=24}^{39} \frac{e^{-30} 30^y}{y!} = 0.8391$

Normal Approximation

Normal approximate Chi-Square

$Y \sim \chi^2_r \xrightarrow{r \text{ large}} \frac{Y - r}{\sqrt{2r}} \sim N(0,1)$

Example : $Y \sim \chi^2_{80}$

$\Pr(60.39 \le Y \le 96.58) \approx 0.9051 - (1 - 0.9394) = 0.8445$

Note : no continuity correction is need as Chi-Square is continuous

From Chi-Square distribution table,

$\Pr(60.39 \le Y \le 96.58) = 0.9 - 0.05 = 0.85$