# Probability
# for Machine Learning

**Industrial AI Lab.**

# Random Variable (= r.v.)

- (Rough) Definition: Variable with a probability

- Probability that $x = a$

$$\triangleq \; P_X(x=a) \; = \; P(x=a) \;\; \Longrightarrow \;\; \begin{cases} 1) \; P(x=a) \geq 0 \\ 2) \; \sum\limits_{\text{all}} P(x) = 1 \end{cases}$$

- $\begin{cases} \text{continuous r.v.} & \text{if } x \text{ is continuous} \\ \text{discrete r.v.} & \text{if } x \text{ is discrete} \end{cases}$

# Random Variable (= r.v.)

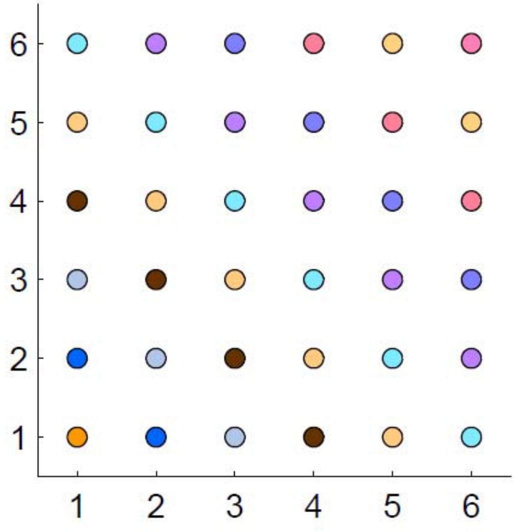
Example

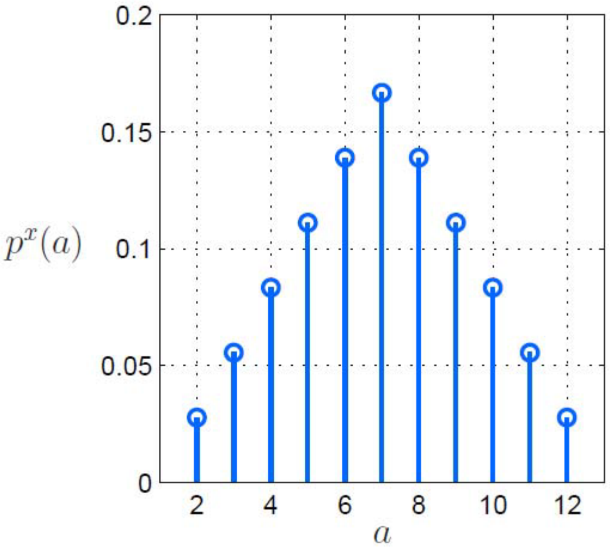
- $x$ : die outcome

$$P(X=1) = P(X=2) = \;\cdots\; = P(X=6) = \frac{1}{6}$$

- Question

$$y = x_1 + x_2 : \qquad \text{sum of two dice}$$
$$P_Y(y=5) = ?$$





# Random Variable (= r.v.)

Expectation = mean

$$E[x] = \begin{cases} \sum_x xP(x) & \text{discrete} \\ \int_x xP(x)dx & \text{continuous} \end{cases}$$

Example

$$\text{Sample mean} \quad E[x] = \sum_x x \cdot \frac{1}{m} \quad (\because \text{ uniform distribution assumed})$$

$$\text{Variance} \quad \text{var}[x] = E\left[(x - E[x])^2\right] \text{ : mean square deviation from mean}$$

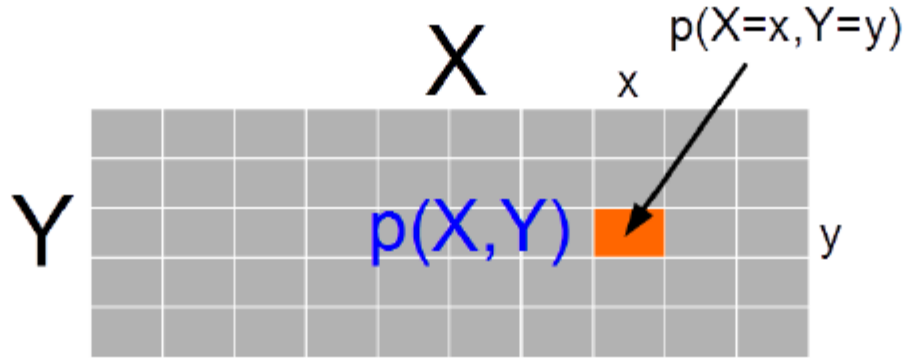
# Random Vectors (multivariate R.V.)

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad n \text{ random variables}$$

# Joint Density Probability

- Joint density probability models probability of co-occurrence of many r.v.

$$P_{X_1,\cdots,X_n}(X_1 = x_1, \cdots, X_n = x_n)$$

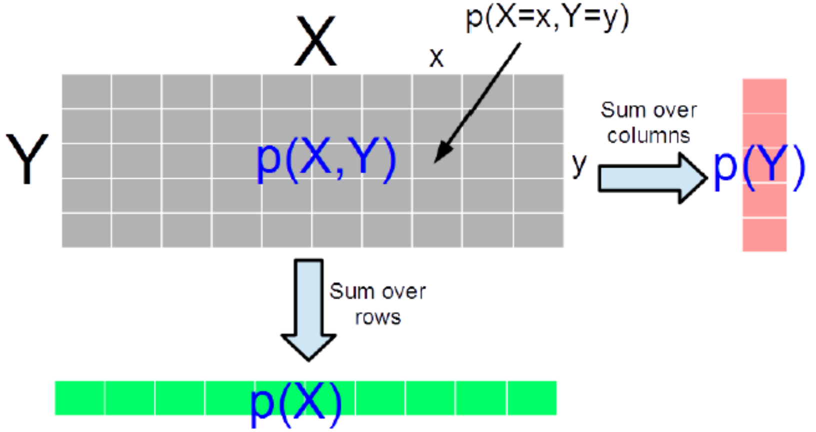


# Marginal Density Probability

$$P_{X_1}(X_1 = x_1)$$
$$\vdots$$
$$P_{X_n}(X_n = x_n)$$

- For two r.v.

$$P(X) = \sum_y P(X, Y = y)$$
$$P(Y) = \sum_x P(X = x, Y)$$

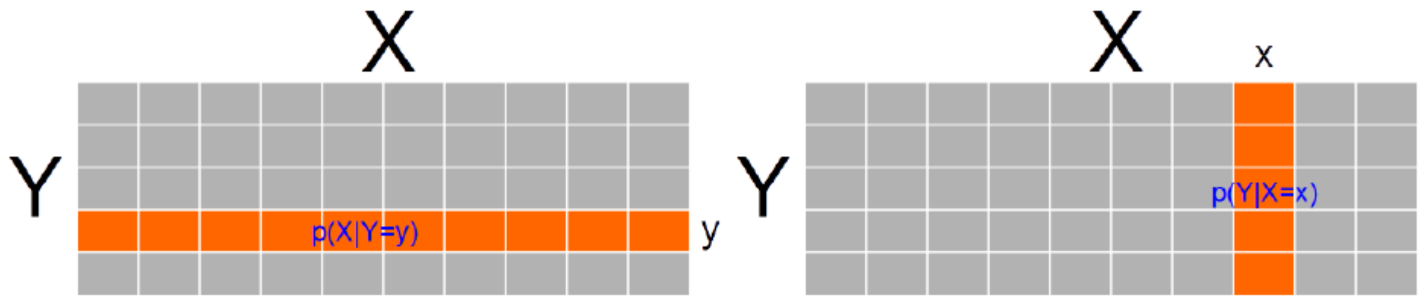


# Conditional Probability

- Probability of one event when we know the outcome of the other

$$P_{X_1|X_2}(X_1 = x_1 \mid X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} :$$

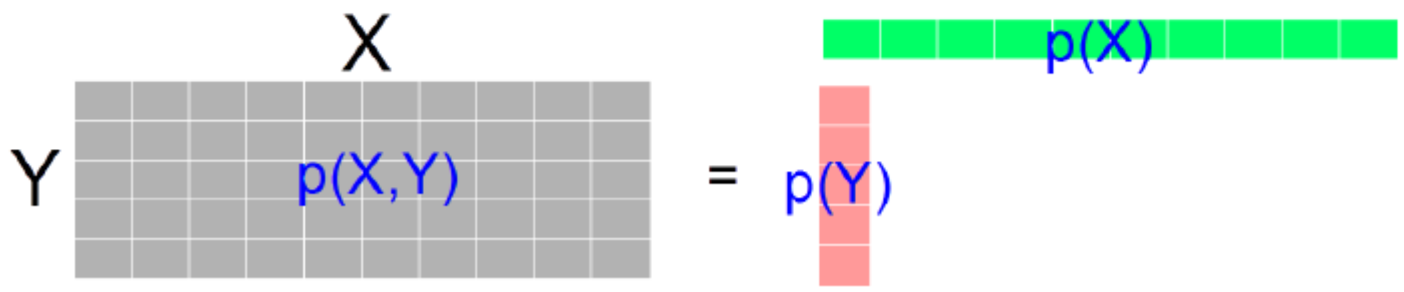
$$\text{Conditional prob. of } x_1 \text{ given } x_2$$



# Conditional Probability

- Independent random variables
  - when one tells nothing about the other

$$P(X_1 = x_1 \mid X_2 = x_2) = P(X_1 = x_1)$$
$$\Updownarrow$$
$$P(X_2 = x_2 \mid X_1 = x_1) = P(X_2 = x_2)$$
$$\Updownarrow$$
$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$$



# Conditional Probability

Example

- four dice $\omega_1, \omega_2, \omega_3, \omega_4$

$$x = \omega_1 + \omega_2 \quad \text{: sum of the first two dice}$$

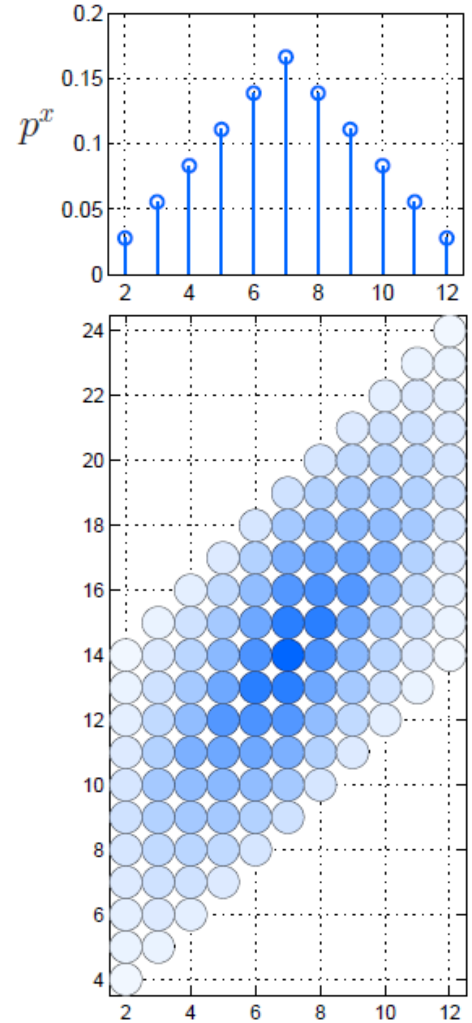
$$y = \omega_1 + \omega_2 + \omega_3 + \omega_4 \quad \text{: sum of all four dice}$$

$$\text{probability of } \begin{bmatrix} x \\ y \end{bmatrix} = ?$$

# Conditional Probability

- marginal probability

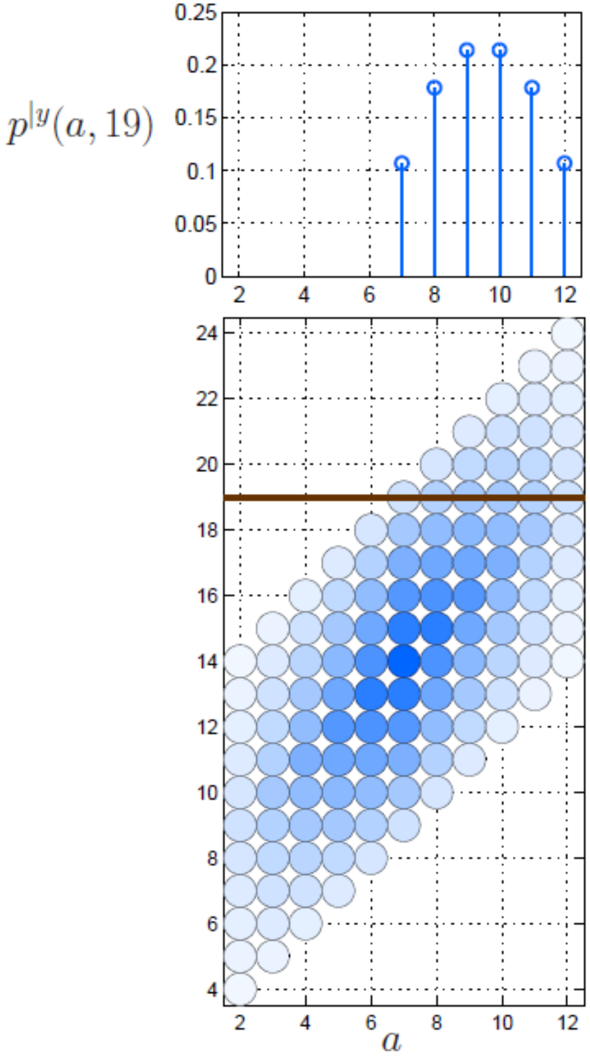
$$P_X(x) = \sum_y P_{XY}(x, y)$$



# Conditional Probability

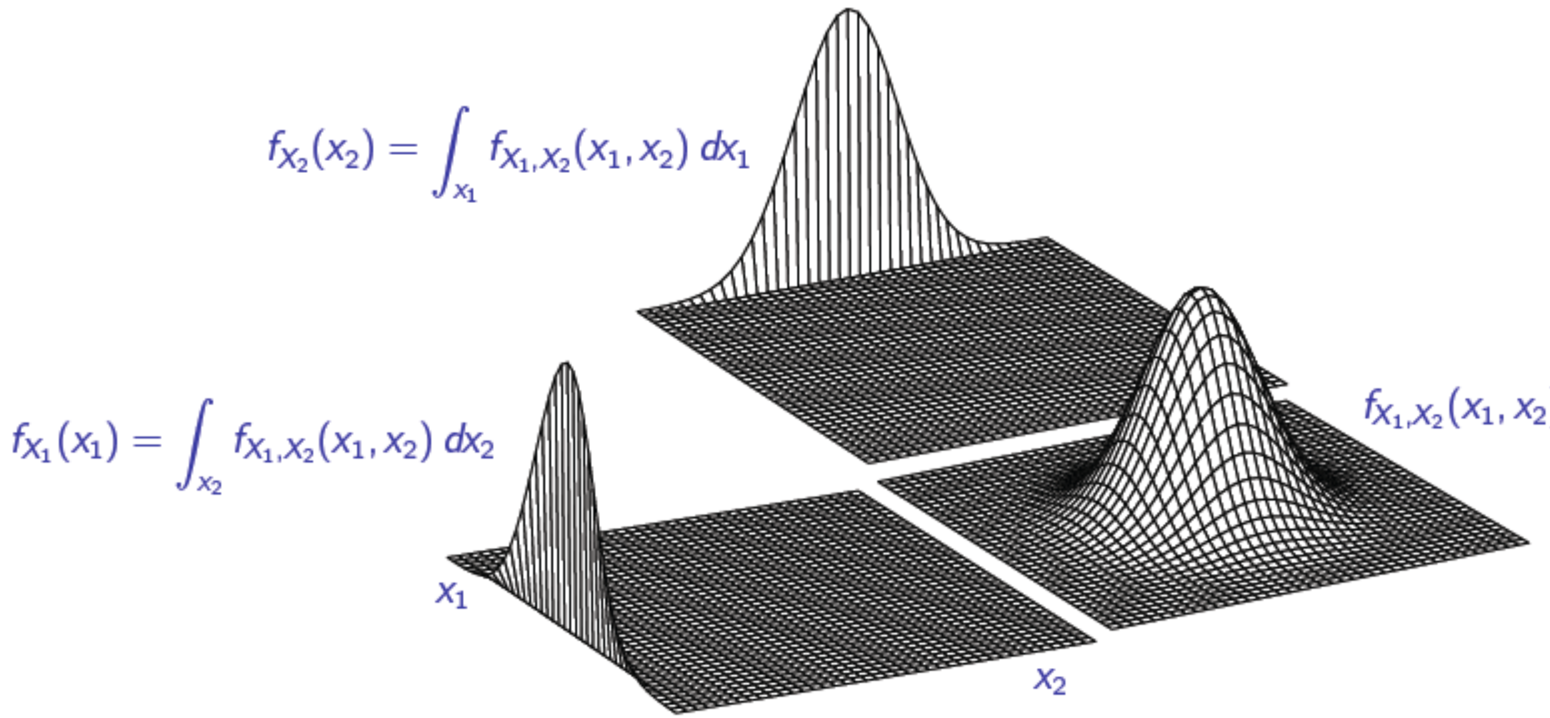
- conditional probability
  - suppose we measured $y = 19$

$$P_{X|Y}(x \mid y = 19) = ?$$



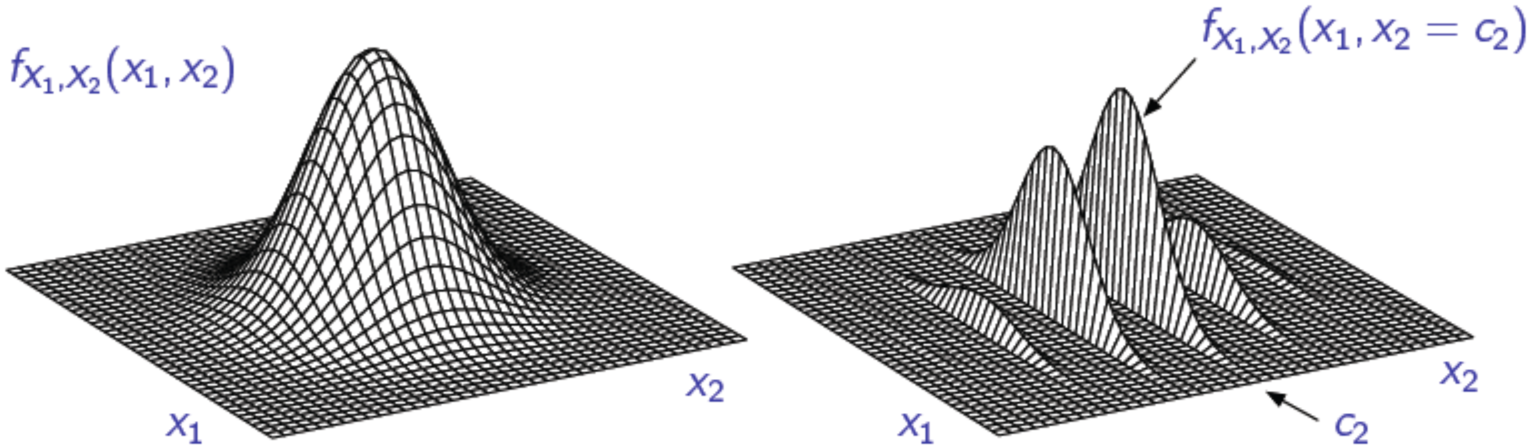
# Conditional Probability

- Pictorial Explanation




- Marginal densities: integrate a continuous joint density (or sum a discrete mass function) over the other variable (by the Law of total probability).

# Conditional Probability




- Conditional density: a “slice" of the joint density, renormalized to integrate to one.

$$f_{X_1|X_2}(x_1|x_2=c_2) = \frac{f_{X_1,X_2}(x_1,x_2=c_2)}{\int_{x_1} f_{X_1,X_2}(x_1,x_2=c_2)\,dx_1}.$$

# Conditional Probability

Example

- Suppose we have three bins, labeled A, B, and C.
- Suppose we have three bins, labeled A, B, and C.

1) We take one ball, what is the probability that it is white? (white = 1)

$$P(X_1 = 1) = \frac{2}{3}$$

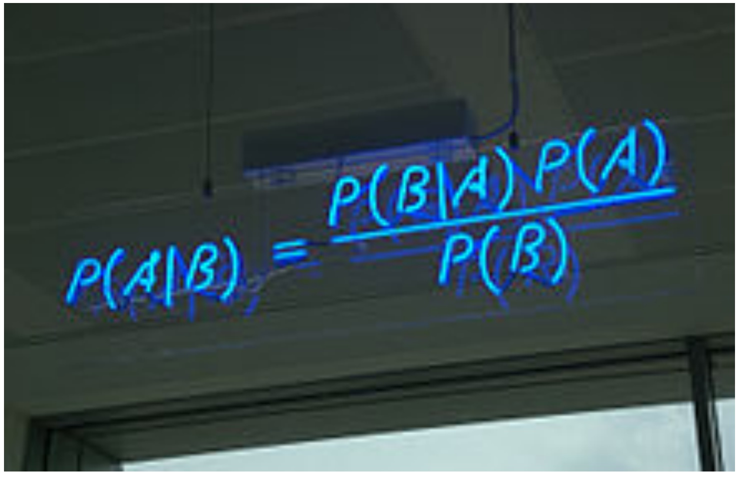
2) When a white ball has been drawn from bin C, what is the probability of drawing a white ball from bin B?

$$P(X_2 = 1 \mid X_1 = 1) = \frac{1}{2}$$

3) When two balls have been drawn from two different bins, what is the probability of drawing two white balls?

$$P(X_1 = 1, X_2 = 1) = P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$$

# Bayes Rule




- Enables us to swap $A$ and $B$ in conditional probability

$$P(X_2, X_1) = P(X_2 \mid X_1)P(X_1) = P(X_1 \mid X_2)P(X_2)$$

$$\therefore \quad P(X_2 \mid X_1) = \frac{P(X_1 \mid X_2)P(X_2)}{P(X_1)}$$

# Bayes Rule

Example

- Suppose that in a group of people, 40% are male and 60% are female.
- 50% of the males are smokers, 30% of the females are smokers.
- Find the probability that a smoker is male

$$
\begin{aligned}
x &= \text{M or F} \qquad & P(x = \mathrm{M}) &= 0.4 \\
y &= \text{S or N} \qquad & P(x = \mathrm{F}) &= 0.6 \\
& & P(y = \mathrm{S} \mid x = \mathrm{M}) &= 0.5 \\
& & P(y = \mathrm{S} \mid x = \mathrm{F}) &= 0.3
\end{aligned}
$$

$$P(x = \mathrm{M} \mid y = \mathrm{S}) = ?$$

- Baye's Rule + conditional probability

$$P(x = \mathrm{M} \mid y = \mathrm{S}) = \frac{P(y = \mathrm{S} \mid x = \mathrm{M})P(x = \mathrm{M})}{P(y = \mathrm{S})} = \frac{0.20}{0.38} \approx 0.53$$

$$
\begin{aligned}
P(y = \mathrm{S}) &= P(y = \mathrm{S} \mid x = \mathrm{M})P(x = \mathrm{M}) + P(y = \mathrm{S} \mid x = \mathrm{F})P(x = \mathrm{F}) \\
&= 0.5 \times 0.4 + 0.3 \times 0.6 = 0.38
\end{aligned}
$$

# Linear Transformation For Single R. V.

$$X \mapsto Y = aX$$

$$\begin{aligned} E[aX] &= aE[X] \\ \mathrm{var}(aX) &= a^2 \mathrm{var}(X) \end{aligned}$$

$$\begin{aligned} \mathrm{var}(X) &= E[(X - E[X])^2] = E[(X - \bar{X})^2] = E[X^2 - 2X\bar{X} + \bar{X}^2] \\ &= E[X^2] - 2E[X\bar{X}] + \bar{X}^2 = E[X^2] - 2E[X]\bar{X} + \bar{X}^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

# Sum of Two Random Variables X and Y

$$Z = X + Y \text{ (still univariate)}$$

$$\begin{aligned}
E[X+Y] &= E[X] + E[Y] \\
\mathrm{var}(X+Y) &= E[(X+Y-E[X+Y])^2] = E[((X-\bar{X}) + (Y-\bar{Y}))^2] \\
&= E[(X-\bar{X})^2] + E[(Y-\bar{Y})^2] + 2E[(X-\bar{X}(Y-\bar{Y})] \\
&= \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)
\end{aligned}$$

$$\begin{aligned}
\mathrm{cov}(X,Y) &= E[(X-\bar{X})(Y-\bar{Y})] = E[XY - X\bar{Y} - \bar{X}Y + \bar{X}\bar{Y}] \\
&= E[XY] - E[X]\bar{Y} - \bar{X}E[Y] - \bar{X}\bar{Y} = E[XY] - E[X]E[Y]
\end{aligned}$$

- Note: quality control in manufacturing process

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$$

# Sum of Two Random Variables X and Y

- **Remark**
  - variance - for univariable
  - covariance - for bivariable
  - Covariance two r.v.

$$\mathrm{cov}(x, y) = E[(x - \mu_x)(y - \mu_y)]$$

  - Covariance matrix for random vectors

$$\mathrm{cov}(X) = E[(X-\mu)(X-\mu)^T] = \begin{bmatrix} \mathrm{cov}(X_1, X_1) & \mathrm{cov}(X_1, X_2) \\ \mathrm{cov}(X_2, X_1) & \mathrm{cov}(X_2, X_2) \end{bmatrix}$$
$$= \begin{bmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_1, X_2) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) \end{bmatrix}$$

  - Moments: provide rough clues on probability distribution

$$\int x^k P_x(x)dx \quad \text{or} \quad \sum x^k P_x(x)dx$$

# Affine Transformation of Random Vectors

$$y = Ax + b$$

$$\begin{aligned} &1.\ \ E[y] = AE[x] + b \\ &2.\ \ \operatorname{cov}(y) = A \operatorname{cov}(x)\, A^T \end{aligned}$$

- IID random variables $\begin{cases} \text{identically distributed} \\ \text{independent} \end{cases}$

- Suppose $x_1, x_2, \cdots, x_m$ are IID with mean $\mu$ and variance $\sigma^2$

$$\text{Let } x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}, \quad \text{then } E[x] = \begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix}, \quad \operatorname{cov}(x) = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{bmatrix}$$

# Affine Transformation of Random Vectors

- Sum of IID random variables (→ single r.v.)

$$S_m = \frac{1}{m}\sum_{i=1}^{m} x_i \implies S_m = Ax \text{ where } A = \frac{1}{m}\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$$

$$E[S_m] = AE[x] = \frac{1}{m}\begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} \mu \\ \vdots \\ \mu \end{bmatrix} = \frac{1}{m} m\mu = \mu$$

$$\text{var}(S_m) = A\,\text{cov}(x)\,A^T = A\begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{bmatrix} A^T = \frac{\sigma^2}{m}$$

- Reduce the variance by a factor of $m \implies$ Law of large numbers or Central limit theorem

$$\bar{x} \longrightarrow N\left(\mu, \left(\frac{\sigma}{\sqrt{m}}\right)^2\right)$$