Statistics and Probability

Agenda

- Elements of probability
- Random variable
- Two and multiple Random variable

Elements of probability

- Sample space Ω: set of all possible outcomes of a random experiment
  {HH, HT, TH, TT}

  Event $A \subseteq \Omega$: subset of Ω {HH, HT} (event A: observe at leave one head)
- Event space $\mathcal{F}$
- Probability measure: $P: \mathcal{F} \rightarrow R$

$$P(A) \geq 0 \text{ and } P(A) \leq 1$$

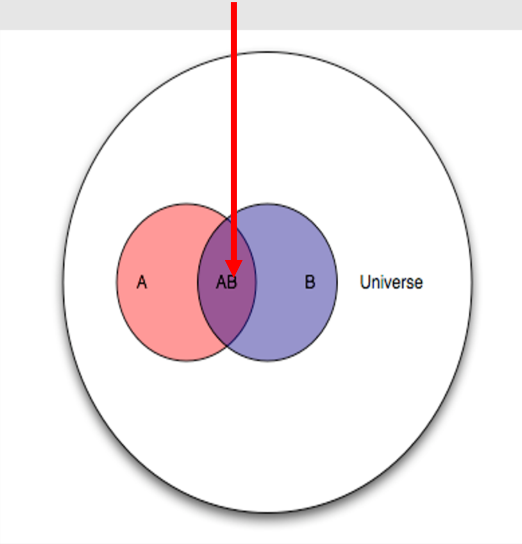
$$P(\Omega) = 1$$

If $A_1, A_2, \ldots$ disjoint set of events $\left(A_i \cap \mathrm{A}_j = \emptyset \; when \; i \neq j\right)$, then $P(\bigcup_i A_i) = \sum_i P(A_i)$

Joint probability

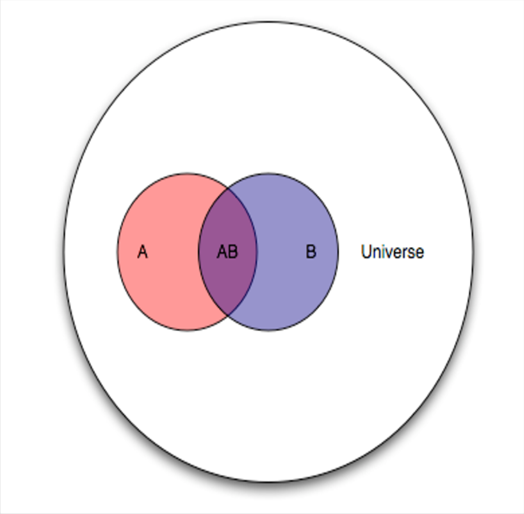
Joint events: the intersection of two events.

$P(A \cap B) = P(AB)$

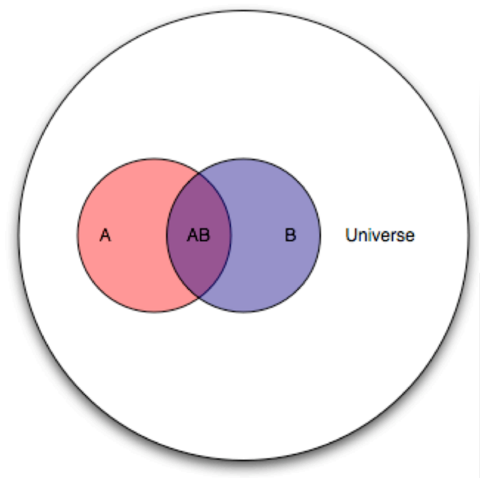


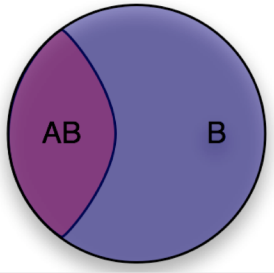
The union of two events

$P(A \cup B) = P(A) + P(B) - P(AB)$



Conditional Probability and Independence





- Let B be any event such that $P(B) \neq 0$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Independence: $A \perp$ B if and only if $P(A \cap B) =$ P(A)P(B)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

Random Variable

Random Variable

Random Variable: outcome of a random experiment.

$$X : \Omega \to R$$

E.g. Flipping a fair coin 10 times, the number of heads: X

$$Val(X) \in \{0, 1, 2, \dots 10\}$$

$$P(X = x)$$

CDF and PDF

CDF

- Cumulative distribution function $F_X(x) = P(X \leq x)$

$$F_X: R \to [0,1]$$

Discrete vs. Continuous Random Variable

Discrete random variable:

$Val(X)$ countable

$P(X = k)$

<u>Probability Mass Function</u>:

$$p_X : Val(X) \to [0,1]$$

$$p_X(x) = P(X = x)$$

$$\sum_{x \in Val(X)} p_X(x) = 1$$

Continuous random variable:

$Val(X)$ uncountable

$P(a \leq X \leq b)$

<u>Probability density Function</u>:

$$f_X : R \to R$$

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$f_X(x) \neq P(X = x)$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$f_X(x) \geq 0$$

Expected Value and Variance

Expected Value:

X is discrete RV with PMF p_X

$$E[X] = \sum_{X \in Val(X)} x \, p_X(x)$$

X is continuous RV with PDF f_X

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

Variance:

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Expected Value and Variance

Expected Value:

X is discrete RV with PMF p_X

$$E[g(X)] = \sum_{X \in Val(X)} g(x)\, p_X(x)$$

X is continuous RV with PDF f_X

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Variance:

$$Var[g(X)] = E(g(X) - E[g(X)])^2 = E[g(X)]^2 - (E[g(X)])^2$$

Expected Value and Variance

Expectation:

$E(a) = a$ for any constant a

$E[aX] = aE[X]$

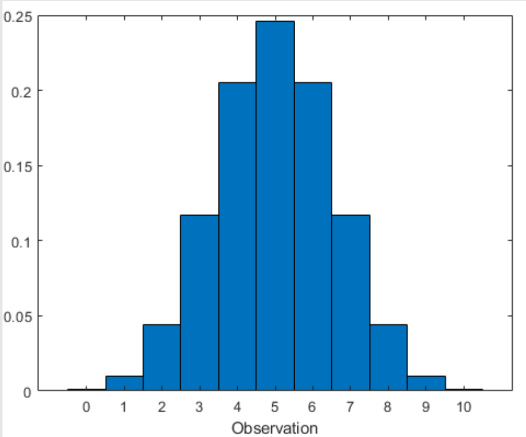
$E\left[f(X) + g(X)\right] = E[f(X)] + E[g(X)]$

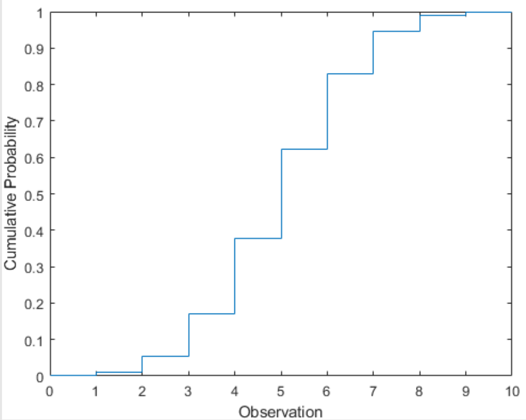
Variance

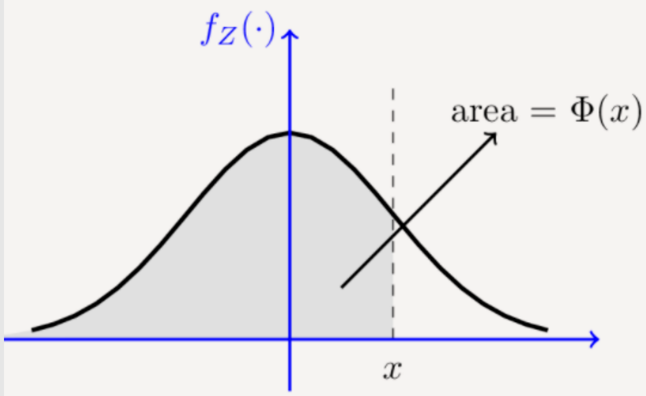
$Var[a] = 0$ for any constant a

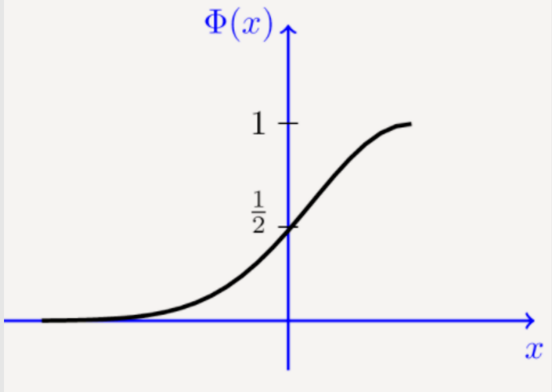
$Var[aX] = a^2 Var[X]$

PDF/PMF and CDF







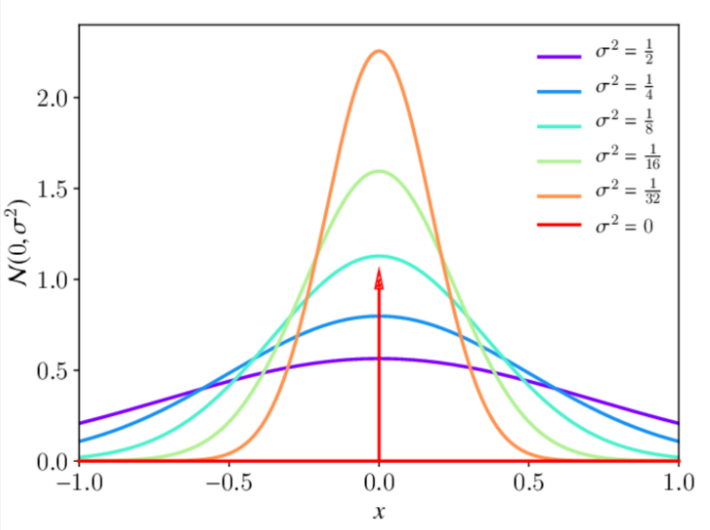


Continuous: Normal Distribution

$X \sim N(\mu, \sigma^2)$

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

$-\infty < x < \infty, \sigma > 0$



<u>Mean and Variance</u>

$$E[X] = \int_{-\infty}^{\infty} xf(x|\mu,\sigma^2)dx = \mu$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x|\mu,\sigma^2)dx = \sigma^2 + \mu^2$$

$$Var(X) = E[X^2] - (E[X])^2 = \sigma^2$$

Continuous: Normal/Gaussian Distribution

$X \sim N(\mu, \sigma^2)$

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

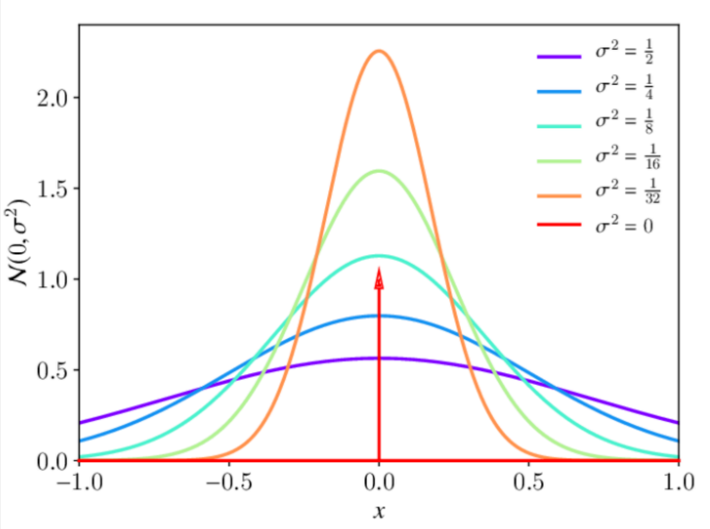
$-\infty < x < \infty, \sigma > 0$

μ: mean determines location

σ^2: variance determines the dispersion of the distribution

<u>Standardization</u>

$$X \sim N(\mu, \sigma^2) \rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$



What if the parameters are unknown

Suppose we have a data set $\{x_1, x_2, \ldots, x_N\}$ drawn independently from random variable X, suppose X follows Gaussian distribution $X \sim N(\mu, \sigma^2)$

$x_1, x_2, \ldots, x_N$ are iid/ independent and identically distributed

Likelihood: $p(x; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$

Log-likelihood: $\log p(x; \mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2 - \frac{N}{2} log\sigma^2 - \frac{N}{2}\log(2\pi)$

Given the data, maximizing the likelihood w.r.t. the parameters

What if the parameters are unknown

Suppose we have a data set $\{x_1, x_2, \ldots, x_N\}$ drawn independently from random variable $X \sim N(\mu, \sigma^2)$

Log-likelihood: $\log p(x; \mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2 - \frac{N}{2} log\sigma^2 - \frac{N}{2}\log(2\pi)$

Maximizing the likelihood w.r.t. the parameters, we get the maximum likelihood estimation (MLE)of the parameters

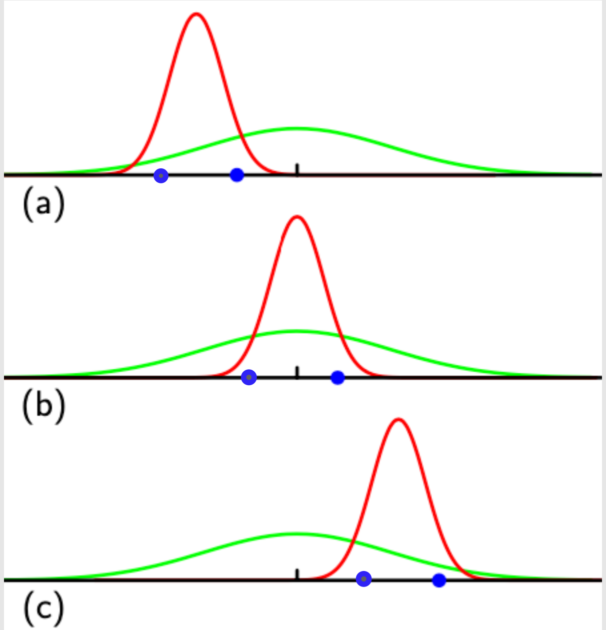
$\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$

$$\sigma_{ML}^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}(x_i - \mu_{ML})^2$$

What if the parameters are unknown

Maximum likelihood estimators for the parameters are random variables too!

$\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i,\ \ \sigma_{ML}^2 = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}(x_i - \mu_{ML})^2$



So we can calculate it's expectations

$E[\mu_{ML}] = \mu, E[\sigma_{ML}^2] = \frac{N-1}{N}\sigma^2$

Adjust for the bias for σ_{ML}^2

$\tilde{\sigma}^2 = \frac{N}{\mathrm{N}-1}\sigma_{ML}^2 = \frac{1}{\mathrm{N}-1}\sum_{i=1}^{N}(x_i - \mu_{ML})^2$

Note: when N→ ∞, the bias of MLE becomes less significant

Discrete: Bernoulli and Binomial

Bernoulli(p)

PMF: $p_X(x) = p^x(1-p)^{1-x} = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$

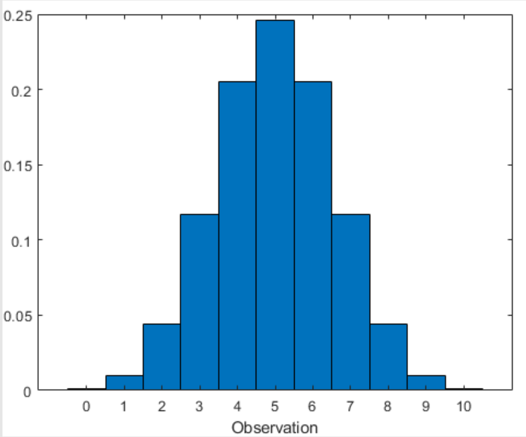
$E[X] = p, Var[X] = p(1-p)$

Binomial(n, p)

number of successes in n independent Bernoulli experiments

PMF: $p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \dots n$

$E[X] = np, Var[X] = np(1-p)$



What if the parameters are unknown

Bernoulli(p)

PMF: $p_X(x) = p^x(1-p)^{1-x} = \begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0 \end{cases}$

$E[X] = p, Var[X] = p(1-p)$

Log–likelihood: $\log P(x;p\ ) = \sum_{i=1}^{N} log\ p_X(x_i;p)$

$$= \sum_{i=1}^{N}\{x_i \log p + (1-x_i)\log(1-p)\}$$

Maximum likelihood estimator for p

$$p_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Law of large numbers

The strong law of large numbers

Let X_1 , X_2 , . . . be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then, with probability 1,

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

Example Distributions

Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p, & \text{if } x = 1 \\ 1 - p, & \text{if } x = 0. \end{cases}$	p	$p(1-p)$
$Binomial(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, ..., n$	np	$np(1-p)$
$Geometric(p)$	$p(1-p)^{k-1}$ for $k = 1, 2, ...$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$\frac{e^{-\lambda}\lambda^k}{k!}$ for $k = 0, 1, ...$	λ	λ
$Uniform(a, b)$	$\frac{1}{b-a}$ for all $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Gaussian(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for all $x \in (-\infty, \infty)$	μ	σ^2
$Exponential(\lambda)$	$\lambda e^{-\lambda x}$ for all $x \geq 0, \lambda \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Two/Multiple Random Variable

Two Random Variables

Bivariate CDF

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

Bivariate PMF

$$p_{XY}(x, y) = P(X = x, Y = y)$$

Marginal PMF

$$p_x(x) = \sum_y p_{XY}(x, y)$$

Bivariate PDF

$$p_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

Marginal PDF

$$p_x(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

Bayes' Theorem

Given the conditional probability of an event P(x|y)

Want to find the "reverse" conditional probability P(y|x)

$$P(y|x) = \frac{P(x,y)}{P(x)} = \frac{P(x|y)P(y)}{P(x)}$$

where $P(x) = \sum_{y'} P(x|y')P(y')$

X and Y are continuous

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{f(x|y)f(y)}{f(x)}$$

where $f(x) = \int f(x|y')f(y')dy'$

Example for Bayes Rule

You randomly choose a treasure chest to open, and then randomly choose a coin from that treasure chest. If the coin you choose is gold, then what is the probability that you choose chest A?

(a) 1/3 (b) 2/3 (c) 1 (d) None

Independence

- Two events are independent if $P(A \cap B)$ = P(A)P(B)

- Two random variables X and Y are independent if:

$$p_{XY}(x, y) = p_X(x)p_Y(y)$$
$$p_{Y|X}(x, y) = p_Y(Y)$$

- For continuous random variables:

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Example for independent variables

Spin a spinner numbered 1 to 7, and toss a coin. What is the probability of getting an odd. number on the spinner and a tail on the coin?

$$p_{XY}(x, y) = p_X(x)p_Y(y) = \frac{1}{2} \times \frac{4}{7} = \frac{2}{7}$$

Expectation

- X,Y: Two continuous random variables

$$E[g(X,Y)] = \int\int g(x,y)f_{XY}(x,y)dxdy$$

- $E[g(X,Y) + f(X,Y)] = E[g(X,Y)] + E[f(X,Y)]$
- If X, Y are independent, $f_{XY}(x,y) = f_X(x)f_Y(y)$, we have

$$E[XY] = E[X]E[Y]$$

Covariance of two random variables

$$Cov[XY] = E[(X - E[X])(Y - E[Y])]$$
$$= E(XY) - E(X)E(Y)$$

- $Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$
- If X and Y are independent then $Cov[XY] = 0$
  $Var[X + Y] = Var[X] + Var[Y]$

Correlation of two random variables

Covariance: $Cov[XY] = E[(X - E[X])(X - E[X])] = E(XY) - E(X)E(Y)$

Pearson's correlation(normalized covariance)

$$\rho_{X,Y} = corr(X,Y) = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

Measures the linear correlation between two features

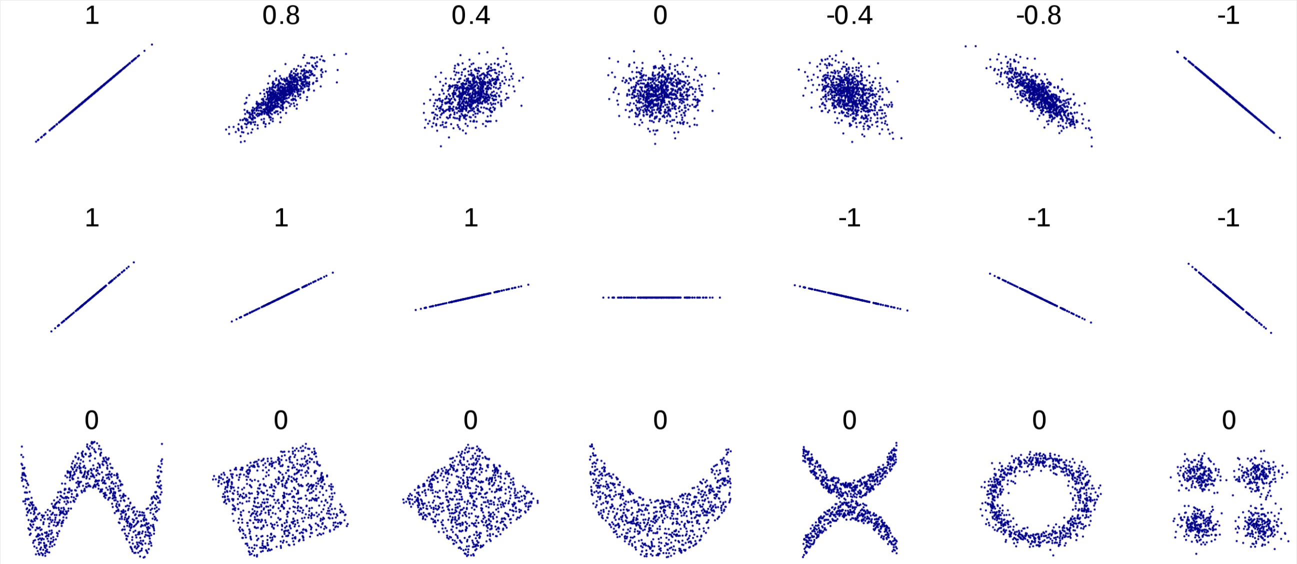
What's the possible range for Pearson's correlation?

$$-1 \leq \rho_{X,Y} \leq 1$$

Correlation of two random variables

Pearson's correlation(normalized covariance)$\rho_{X,Y} = corr(X,Y) = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$

Measures the linear correlation between two features



When the true distribution is unknown

Draw a random sample $x_1, x_2, \ldots x_N$ from the random variable X.
Draw a random sample $y_1, y_2, \ldots y_N$ from the random variable Y.

Sample Mean: $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$

Sample variance: $\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$

Sample covariance: $\hat{\sigma}_{XY}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$

Multiple Random Variables

Random Variable $X_1, X_2, \ldots X_n$, the joint distribution

$$F_{X_1,\ldots X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots X_n \le x_n)$$

Joint probability density function

$$f_{X_1,\ldots X_n}(x_1, \ldots, x_n) = \frac{\partial^n F_{X_1,\ldots X_n}(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

Chain rule:

$$f(x_1, \ldots, x_n) = f(x_n | x_1, \ldots, x_{n-1}) f(x_1, \ldots, x_{n-1})$$

$$= \cdots = f(x_1) \prod_{i=2}^{n} f(x_i | x_1, \ldots, x_{i-1})$$

If $X_1, X_2, \ldots X_n$ are independent $f(x_1, \ldots, x_n) = f(x_1)\, f(x_2) \ldots f(x_n)$

Multiple Random Variables

Random vector:

$$X = \begin{bmatrix} X_1 \\ \dots \\ X_n \end{bmatrix}$$

Expectation

$$E(X) = [E(X_1), \dots E(X_n)]^T$$

Covariance matrix:

$$\Sigma = \begin{bmatrix} Var[X_1] & \dots & Cov[X_1, X_n] \\ \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & \dots & Var[X_n] \end{bmatrix} = E(X - E[X])(X - E[X])^T$$

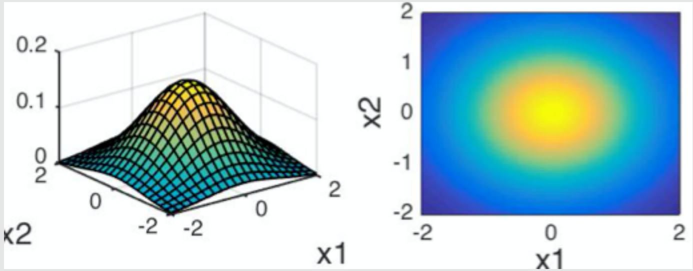
Multivariant Gaussian (Normal) Distribution

x $\in R^n$. Model $p(x_1), p(x_2)$ ... at the same time. Parameters :μ $\in R^n$, $\Sigma \in R^{n\times n}$ (covariance matrix)

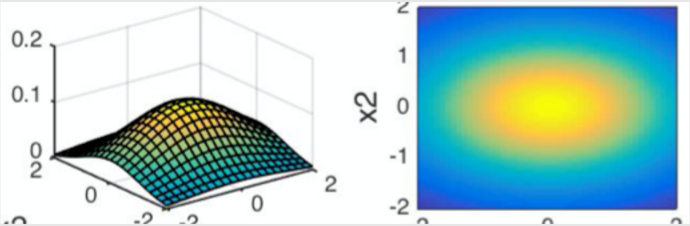
$$p(x, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Multivariant Gaussian Distribution

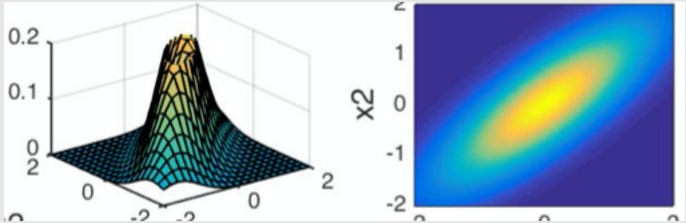
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\mu = [0,0]^T$$



$$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\mu = [0,0]^T$$



$$\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$
$$\mu = [0,0]^T$$



Marginal and Conditional Gaussian Distributions(Optional)

$X \sim N(\mu, \Sigma)$ and we partition X into two disjoint subsets

$$X = \begin{bmatrix} X_a \\ X_b \end{bmatrix}$$

Similarly we partition the parameters

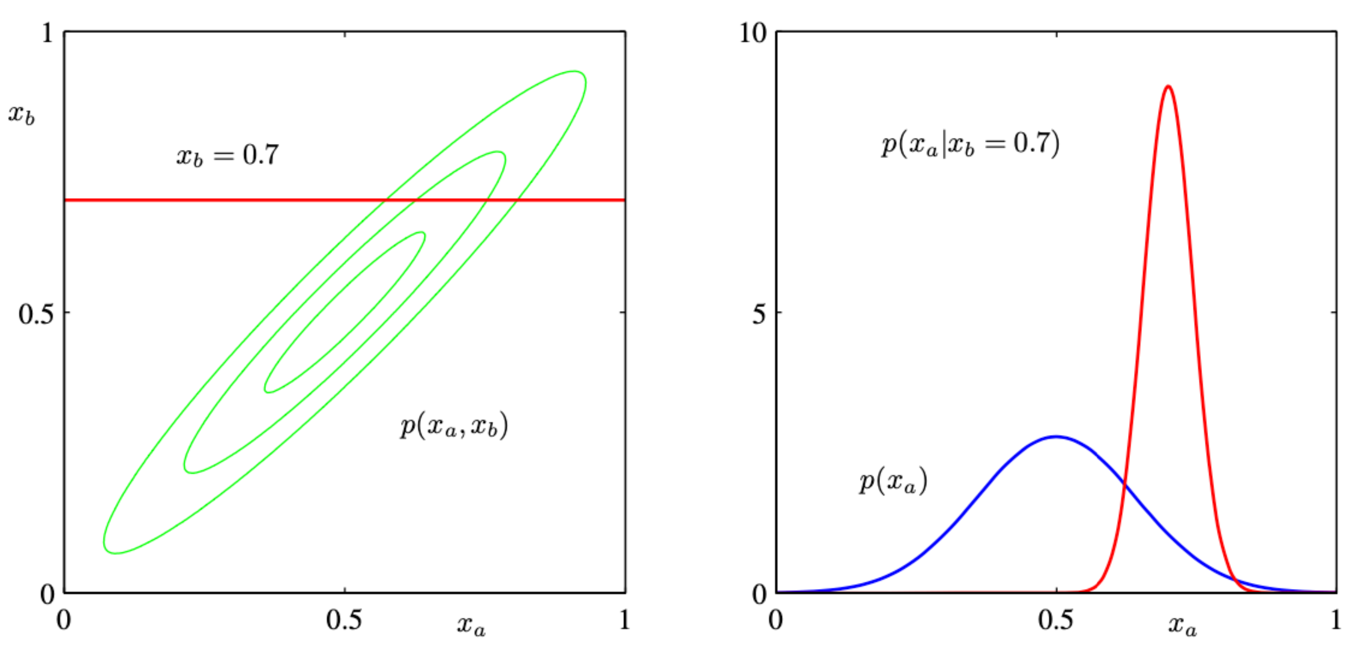
$$\mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}, \Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{bmatrix}$$

Then the marginal distribution for X_a is $X_a \sim N(\mu_a, \Sigma_{aa})$

The conditional distribution $X_a | X_b \sim N\left(\mu_{a|b}, \Lambda_{aa}^{-1}\right)$

where $\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (X_b - \mu_b)$

Marginal and Conditional Gaussian Distributions(Optional)



Conditional Probability and Expectation

Recall that $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Given two random variables X, Y, we have

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$E(X | Y = y) = \sum_x x \frac{P(X = x, Y = y)}{P(Y = y)}$$

Conditional Probability and Expectation

$E(X|Y=y)$ is a function of Y
$E(X|Y)$ is a random variable!

Law of total expectation
Let X, Y be random variables with the same probability space then
$E[X] = E[E(X|Y)]$

Conditional Probability and Expectation (Optional)

Law of total expectation

Let X, Y be random variables with the same probability space then $E[X] = E[E(X|Y)]$

$$
\begin{aligned}
E[E(X|Y)] &= E[\textstyle\sum_x xP(X = x|Y)] \\
&= \textstyle\sum_y (\sum_x xP(X = x|Y))P(Y = y) \\
&= \textstyle\sum_y \sum_x xP(X = x, Y = y) \\
&= \textstyle\sum_x x\left(\sum_y P(X = x, Y = y)\right) \\
&= \textstyle\sum_x xP(X = x) \\
&= E[X]
\end{aligned}
$$