Review of Probability Theory

Zahra Koochak and Jeremy Irvin

Elements of Probability

Sample Space Ω

$\{HH, HT, TH, TT\}$

Event $A \subseteq \Omega$

$\{HH, HT\}$, Ω

Event Space $\mathcal{F}$

Probability Measure $P : \mathcal{F} \to \mathbb{R}$

$P(A) \geq 0 \quad \forall A \in \mathcal{F}$

$P(\Omega) = 1$

If $A_1, A_2, ...$ disjoint set of events ($A_i \cap A_j = \emptyset$ when $i \neq j$), then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Conditional Probability and Independence

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

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

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

$A \perp B$ if and only if $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$

Random Variables (RV)

sequence of heads and tails

$$\omega_0 = HHHTHTTHTT$$

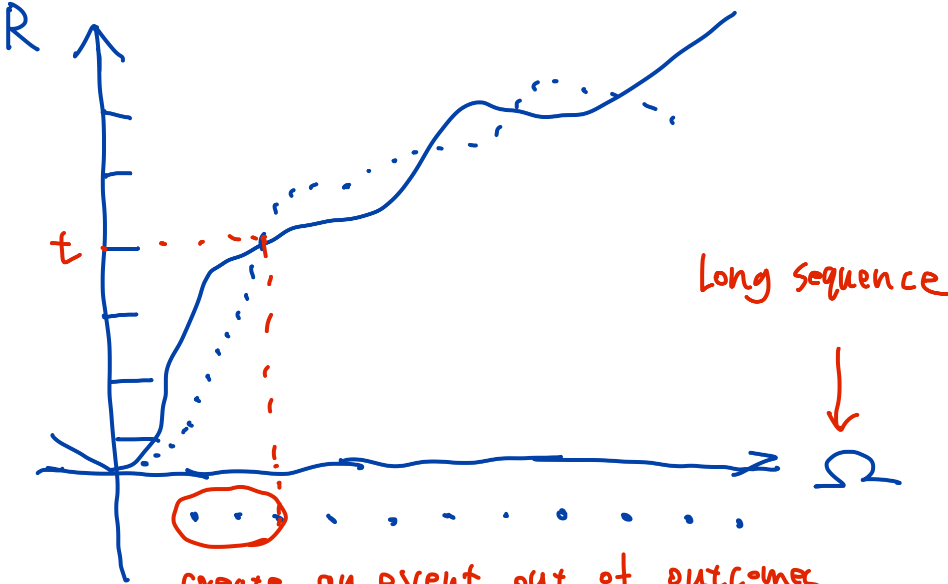
A **RV** is $X : \Omega \to \mathbb{R}$ $\boxed{\text{function}}$ map outcome to real value

# of heads: $X(\omega_0) = 5$

# of tosses until tails: $X(\omega_0) = 4$

$Val(X) := X(\Omega)$

$Val(X) = \{0, 1, ..., 10\}$



create an event out of outcomes and measure the probability on that event

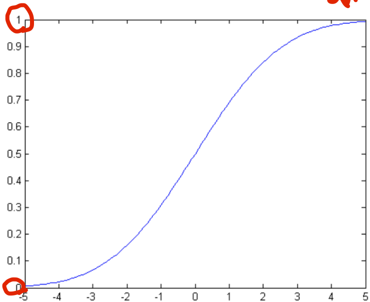
Cumulative Distribution Function (CDF)

$$F_X : \mathbb{R} \to [0, 1]$$

RV of w ≤ X

$$F_X(x) = P(X \leq x) := P(\{\omega | \boxed{X(\omega) \leq x}\})$$

an outcome



Discrete vs. Continuous RV

Discrete RV $Val(X)$ countable

$P(X = k) := P(\{\omega | X(\omega) = k\})$

Probability Mass Function (PMF)

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

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

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

Continuous RV $Val(X)$ uncountable

$P(a \leq X \leq b) := P(\{\omega | a \leq X(\omega) \leq b\})$

Probability Density Function (PDF)

$f_X : \mathbb{R} \to \mathbb{R}$

$$f_X(x) := \frac{d}{dx} F_X(x)$$

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

$$\int_{-\infty}^{\infty} \underbrace{f_X(x)dx}_{P(x \leq X \leq x+dx)} = 1$$

Expected Value and Variance

$g : \mathbb{R} \to \mathbb{R}$

Expected Value

Let X be a discrete RV with PMF p_X.

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

Let X be a continuous RV with PDF f_X.

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

Variance

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

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(λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$ 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(μ, σ^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(λ)	$\lambda e^{-\lambda x}$ for all $x \geq 0, \lambda \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

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

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

Marginal PDF

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

Bayes' Theorem

- Given the conditional probability of an event $P(x|y)$
- Want to find the "reverse" conditional probability, $P(y|x)$

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

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

X and Y are continuous

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

where: $f(x) = \int_{y' \in value\ y} f(x|y')f(y')dy'$

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?

$$a)\frac{1}{3} \qquad b)\frac{2}{3} \qquad c)1 \qquad d)None$$

Independence

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:

$$p_{XY}(x, y) \to f_{XY}(x, y)$$

Example for independent random 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
- g , R2 $\rightarrow$ R : A function of X and Y

$$E(g(x,y)) = \int_{x \in Val(x)} \int_{y \in Val(y)} g(x,y) f_{XY}(x,y) dxdy$$

Example

$g(x,y) = 3x, f_{x,y} = 4xy, \;\; 0 < x < 1, \;\; 0 < y < 1$
$E(g(x,y)) = \int_0^1 \int_0^1 3x \times 4xy \;\, dxdy$

Covariance of two random variables X and Y

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

If X and Y are independent, then:

$$E(XY) = E(X)E(Y) \rightarrow Cov[x, y] = 0$$

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

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Multivariant Gaussian (Normal) distribution

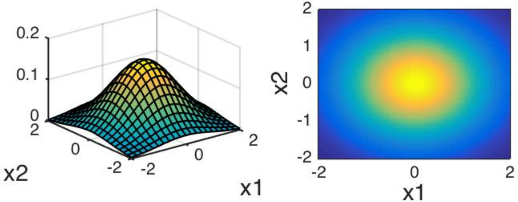
$x \in \mathrm{I\!R^n}$. Model $p(x_1), p(x_2), ...etc$. at the same time. Parameters $:\mu \in \mathrm{I\!R^n}, \Sigma \in \mathrm{I\!R^{n \times n}}(\mathrm{covariancematrix})$

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

Multivariant Gaussian (Normal examples)

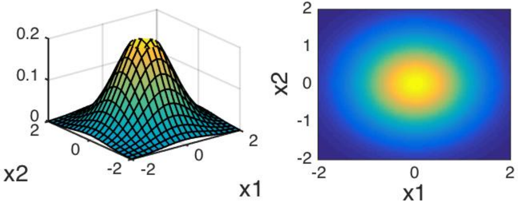
$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



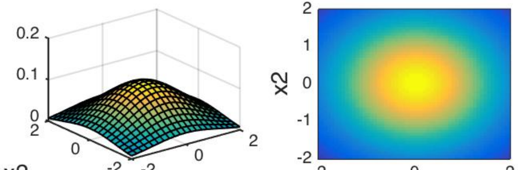
$$\Sigma = \begin{matrix} 0.7 & 0 \\ 0 & 0.7 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



$$\Sigma = \begin{matrix} 1.5 & 0 \\ 0 & 1.5 \end{matrix}$$

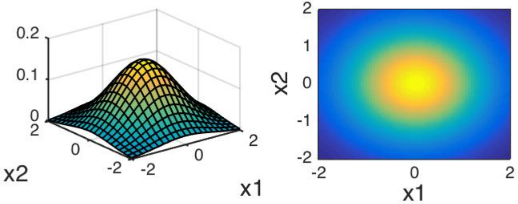
$$\mu = [0 \;\; 0]^T$$



Multivariant Gaussian (Normal examples)

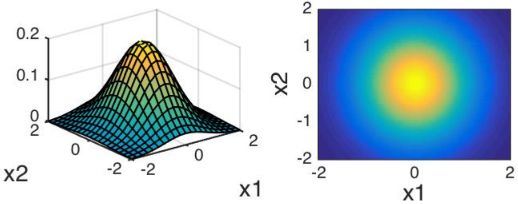
$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



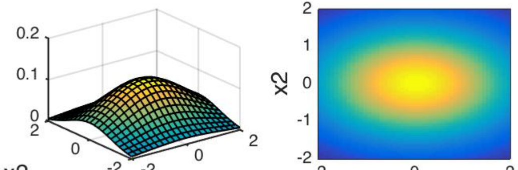
$$\Sigma = \begin{matrix} 0.6 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



$$\Sigma = \begin{matrix} 2 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \;\; 0]^T$$



Multivariant Gaussian (Normal examples)

$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

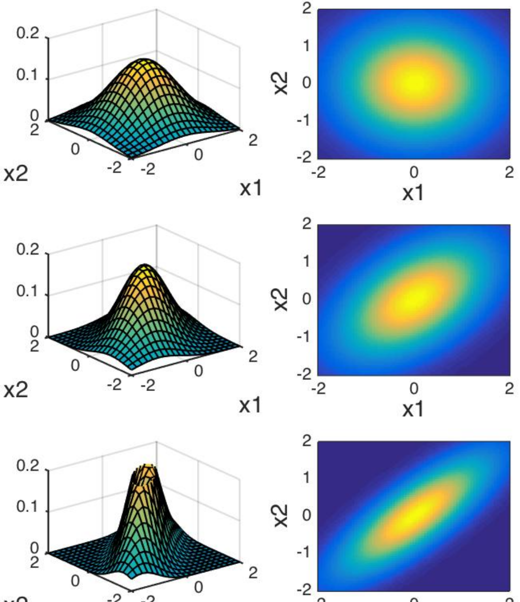
$$\mu = [0 \ \ 0]^T$$

$$\Sigma = \begin{matrix} 1 & 0.5 \\ 0.5 & 1 \end{matrix}$$

$$\mu = [0 \ \ 0]^T$$

$$\Sigma = \begin{matrix} 1 & 0.8 \\ 0.8 & 1 \end{matrix}$$

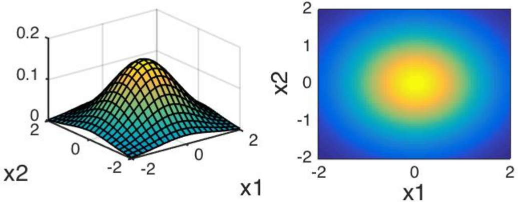
$$\mu = [0 \ \ 0]^T$$



Multivariant Gaussian (Normal examples)

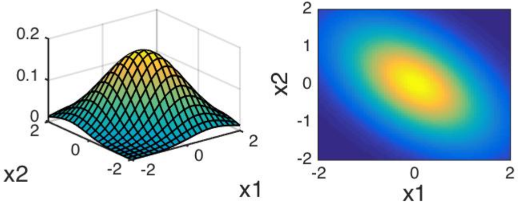
$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \quad 0]^T$$



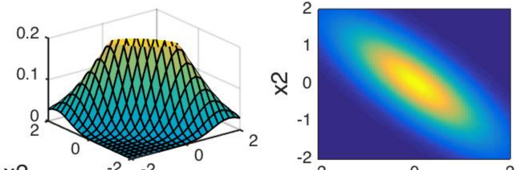
$$\Sigma = \begin{matrix} 1 & -0.5 \\ -0.5 & 1 \end{matrix}$$

$$\mu = [0 \quad 0]^T$$



$$\Sigma = \begin{matrix} 1 & -0.8 \\ -0.8 & 1 \end{matrix}$$

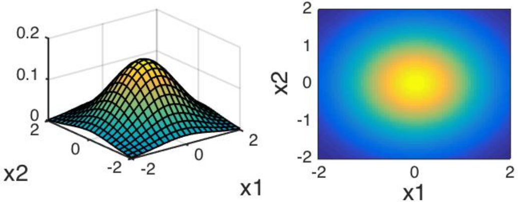
$$\mu = [0 \quad 0]^T$$



Multivariant Gaussian (Normal examples)

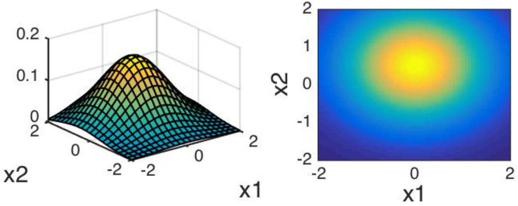
$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \quad 0]^T$$



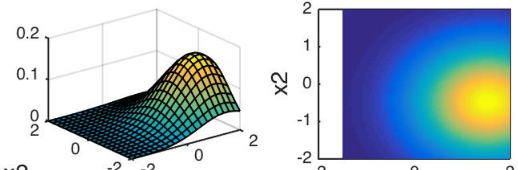
$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [0 \qquad 0.5]^T$$



$$\Sigma = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$$

$$\mu = [1.5 \qquad -0.5]^T$$



Conditional Probability and Expectation

Remember:

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

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

Conditional Probability and Expectation

X,Y are RVs with the same probability space,

we have

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

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

Conditional Probability and Expectation

$\mathbb{E}[X|Y]$

It is actually a random variable

$\mathbb{E}[X|Y](y) = \mathbb{E}[X|Y = y]$ is a function of Y

Law of Total Expectation

Let X, Y be RVs with the same probability space, then

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$

A brief proof of X,Y being discrete and finite

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

More Conditioned Bayes Rule

$$P(a|b,c) = \frac{P(b|a,c)P(a|c)}{P(b|c)}$$

It is actually the same as the Bayes Rule:

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)}$$

with a random variable c that all the probabilities are conditioned on.

More Conditioned Bayes Rule

A proof:

$$\begin{aligned}
\frac{P(b|a,c)P(a|c)}{P(b|c)} &= \frac{P(b,a,c)P(a|c)}{P(b|c)P(a,c)} \\
&= \frac{P(b,a,c)P(a,c)}{P(b|c)P(a,c)P(c)} \\
&= \frac{P(b,a,c)}{P(b|c)P(c)} \\
&= \frac{P(b,a,c)}{P(b,c)} \\
&= P(a|b,c)
\end{aligned}$$

Statistic

Parameters   Inference   observation / Data

Learning / Training / Fitting

μ, Σ ← model ← ML — Traing data $X \in \mathbb{R}^n$

Prediction → Future Data

- Method of momt

Probability

- Maximum Likehood Estimation

how good the model to predict future data

Training Data

(x, y)

$\vdots$

Gaussian Data

$$P(x^{(1)}, \dots, x^{(n)}) = \prod_{i=1}^{n} P(x^{(i)})$$

i到n乘积

$x \in \mathbb{R}^d$, Simple

$X = \{x^{(1)}, \dots, x^{(n)}\}$, $x^{(i)}$ – i^{th} example $\in \mathbb{R}^d$,

independently & identically Distributed

$$P(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right]$$

$$P(x^{(1)}, \dots, x^{(n)}; \mu, \Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |\Sigma^{1/2}|} \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x^{(i)}-\mu)\right]$$

$$L(\mu, \Sigma; X) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x^{(1)}-\mu) \Sigma^{-1} (x^{(i)} - \mu)\right]$$

Likehood function

Probability of Data given parameters

Likelihood of parameters give data