

Review of Probability Theory

Zahra Koochak and Jeremy Irvin

Elements of Probability

Elements of Probability

Sample Space Ω

Elements of Probability

Sample Space Ω

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

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\},$$

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Event Space \mathcal{F}

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Event Space \mathcal{F}

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

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Event Space \mathcal{F}

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

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

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Event Space \mathcal{F}

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

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

$$P(\Omega) = 1$$

Elements of Probability

Sample Space Ω

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

Event $A \subseteq \Omega$

$$\{HH, HT\}, \Omega$$

Event Space \mathcal{F}

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

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

$$P(\Omega) = 1$$

If A_1, A_2, \dots 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

Conditional Probability and Independence

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

Conditional Probability and Independence

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

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

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

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

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

Random Variables (RV)

Random Variables (RV)

$$\omega_0 = HHHTHHTTTT$$

Random Variables (RV)

$$\omega_0 = HHHTHHTTTT$$

A **RV** is $X : \Omega \rightarrow \mathbb{R}$

Random Variables (RV)

$$\omega_0 = HHHTHHTTTT$$

A **RV** is $X : \Omega \rightarrow \mathbb{R}$

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

Random Variables (RV)

$$\omega_0 = HHH THTTHTT$$

A **RV** is $X : \Omega \rightarrow \mathbb{R}$

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

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

Random Variables (RV)

$$\omega_0 = HHHHTHTTHTT$$

A **RV** is $X : \Omega \rightarrow \mathbb{R}$

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

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

$$\text{Val}(X) := X(\Omega)$$

Random Variables (RV)

$$\omega_0 = HHHHTHTTHTT$$

A **RV** is $X : \Omega \rightarrow \mathbb{R}$

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

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

$$\text{Val}(X) := X(\Omega)$$

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

Cumulative Distribution Function (CDF)

Cumulative Distribution Function (CDF)

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

Cumulative Distribution Function (CDF)

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

$$F_X(x) = P(X \leq x)$$

Cumulative Distribution Function (CDF)

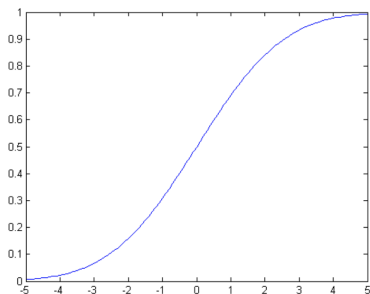
$$F_X : \mathbb{R} \rightarrow [0, 1]$$

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

Cumulative Distribution Function (CDF)

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

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



Discrete vs. Continuous RV

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Continuous RV: $Val(X)$ uncountable

Probability Mass Function (PMF)

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

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Probability Density Function (PDF)

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Probability Density Function (PDF)

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Probability Density Function (PDF)

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

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Probability Density Function (PDF)

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

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

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

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

Discrete vs. Continuous RV

Discrete RV: $Val(X)$ countable

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

Probability Mass Function (PMF)

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

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

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

Continuous RV: $Val(X)$ uncountable

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

Probability Density Function (PDF)

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

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

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

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

Expected Value and Variance

Expected Value and Variance

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

Expected Value and Variance

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

Expected Value

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Let X be a continuous RV with PDF f_X .

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Let X be a continuous RV with PDF f_X .

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

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Let X be a continuous RV with PDF f_X .

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

Variance

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Let X be a continuous RV with PDF f_X .

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

Variance

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

Expected Value and Variance

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

Expected Value

Let X be a discrete RV with PMF p_X .

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

Let X be a continuous RV with PDF f_X .

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

Variance

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

Example Distributions

Distribution	PDF or PMF	Mean	Variance
<i>Bernoulli</i> (p)	$\begin{cases} p, & \text{if } x = 1 \\ 1 - p, & \text{if } x = 0. \end{cases}$	p	$p(1 - p)$
<i>Binomial</i> (n, p)	$\binom{n}{k} p^k (1 - p)^{n-k}$ for $k = 0, 1, \dots, n$	np	$np(1 - p)$
<i>Geometric</i> (p)	$p(1 - p)^{k-1}$ for $k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
<i>Poisson</i> (λ)	$\frac{e^{-\lambda} \lambda^k}{k!}$ for $k = 0, 1, \dots$	λ	λ
<i>Uniform</i> (a, b)	$\frac{1}{b-a}$ for all $x \in (a, b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
<i>Gaussian</i> (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
<i>Exponential</i> (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: $P(x) = \sum_{y' \in \text{value } y} P(x|y')P(y')$

X and Y are continuous

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

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

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) $\frac{1}{3}$

b) $\frac{2}{3}$

c) 1

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:

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

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
- ▶ $g, \mathbb{R}^2 \rightarrow \mathbb{R}$: A function of X and Y

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

Example

$$g(x, y) = 3x, f_{x,y} = 4xy, 0 < x < 1, 0 < y < 1$$

$$E(g(x, y)) = \int_0^1 \int_0^1 3x \times 4xy \, dx dy$$

Covariance of two random variables X and Y

$$\begin{aligned}\text{Cov}[x, y] &= E[(x - E[x])(y - (E[y]))] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

If X and Y are independent, then:

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

$$\text{Var}[X + Y] = [E(X + Y)]^2 - E((X + Y)^2)$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

Multivariate Gaussian (Normal) distribution

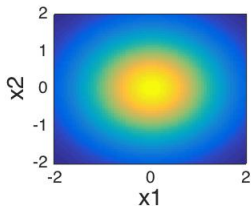
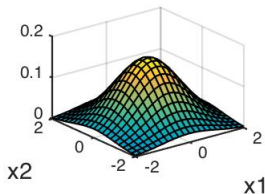
$x \in \mathbb{R}^n$. Model $p(x_1), p(x_2), \dots$ etc. at the same time. Parameters
: $\mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n \times n}$ (covariance matrix)

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

Multivariate Gaussian (Normal examples)

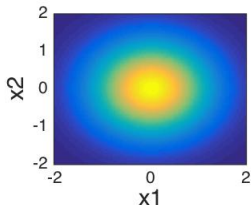
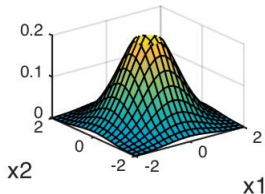
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



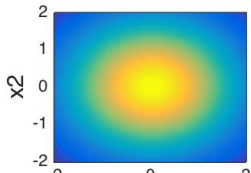
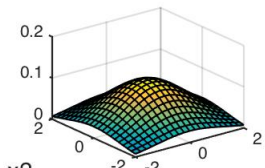
$$\Sigma = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}$$

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



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

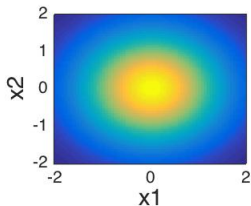
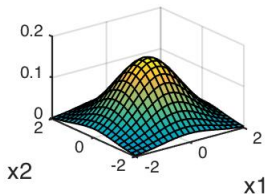
$$\mu = [0 \ 0]^T$$



Multivariate Gaussian (Normal examples)

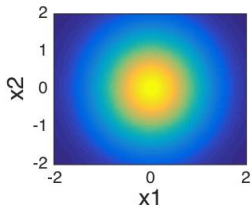
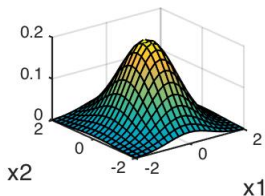
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



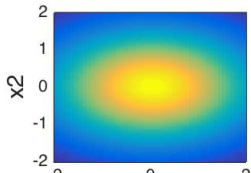
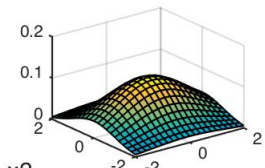
$$\Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix}$$

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



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

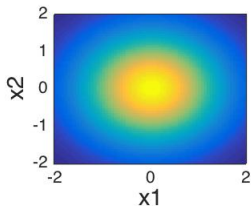
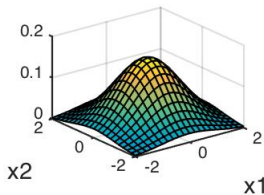
$$\mu = [0 \ 0]^T$$



Multivariate Gaussian (Normal examples)

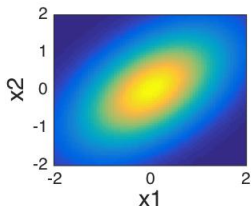
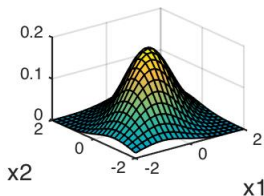
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



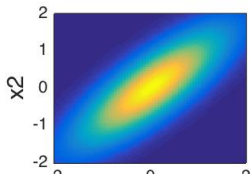
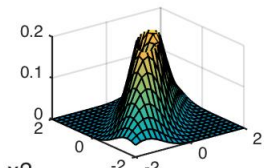
$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

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



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

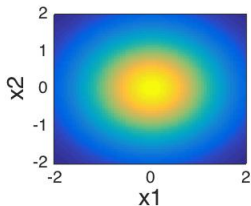
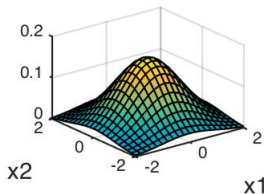
$$\mu = [0 \ 0]^T$$



Multivariate Gaussian (Normal examples)

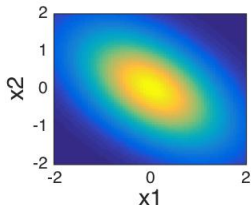
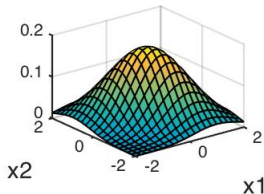
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



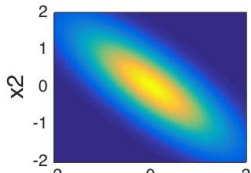
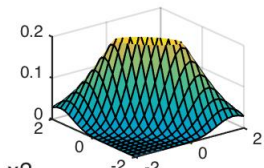
$$\Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

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



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

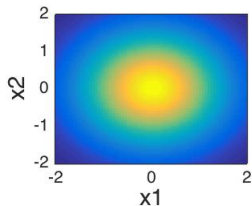
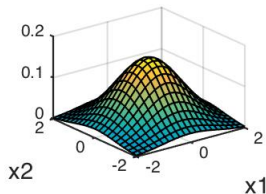
$$\mu = [0 \ 0]^T$$



Multivariate Gaussian (Normal examples)

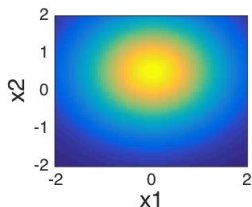
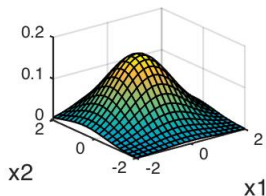
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



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

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



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

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

