

# Assignment Project Exam Help

(CS229 : Probability Theory Review)

Taide Ding, Fereshte Khani  
<https://powcoder.com>

Based on CS229 Review of Probability Theory by Arian Maleki and Tom Do.  
Additional material by Zahra Koochak and Jeremy Irvin.

Add WeChat powcoder

April 9, 2021

N.B.

# Assignment Project Exam Help

This review assumes basic background in probability (events, sample space, probability axioms etc.) and focuses on concepts useful to CS229 and to machine learning in general.

<https://powcoder.com>

Add WeChat powcoder

## Conditional Probability and Bayes' Rule

For any events  $A, B$  such that  $P(B) \neq 0$ , we define:

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

# Assignment Project Exam Help

Let's apply conditional probability to obtain **Bayes' Rule!**

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

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

# Add WeChat powcoder

**Conditioned Bayes' Rule:** given events  $A, B, C$ ,

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

See Appendix for proof :)

## Law of Total Probability

Let  $B_1, \dots, B_n$  be  $n$  disjoint events whose union is the entire sample space. Then, for any event  $A$ ,

**Assignment Project Exam Help**

$$P(A) = \sum_{i=1}^n P(A \cap B_i)$$

$$= \sum_{i=1}^n P(A | B_i) P(B_i)$$

<https://powcoder.com>

We can then write Bayes' Rule as:

**Add WeChat powcoder**

$$P(B_k | A) = \frac{P(B_k) P(A | B_k)}{P(A)}$$

$$= \boxed{\frac{P(B_k) P(A | B_k)}{\sum_{i=1}^n P(A | B_i) P(B_i)}}$$

## Example

Treasure chest **A** holds 100 gold coins. Treasure chest **B** holds 60 gold and 40 silver coins.

Choose a treasure chest uniformly at random, and pick a coin from that chest uniformly at random. If the coin is gold, then what is the probability that you chose chest **A**?<sup>1</sup>

**Solution:**

$$P(A | G) = \frac{P(A)P(G | A)}{P(A)P(G | A) + P(B)P(G | B)}$$
$$= \frac{0.5 \times 1}{0.5 \times 1 + 0.5 \times 0.6}$$

$$= \boxed{0.625}$$

<sup>1</sup>Question based on slides by Koochak & Irvin

# Assignment Project Exam Help

For any  $n$  events  $A_1, \dots, A_n$ , the joint probability can be expressed as a product of conditionals:

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1)P(A_2 | A_1)P(A_3 | A_2 \cap A_1)\dots P(A_n | A_{n-1} \cap A_{n-2} \cap \dots \cap A_1) \end{aligned}$$

Add WeChat powcoder

## Independence

Events  $A, B$  are independent if

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

# Assignment Project Exam Help

We denote this as  $A \perp B$ . From this, we know that if  $A \perp B$ ,

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

**Implication:** If two events are independent, observing one event does not change the probability that the other event occurs.

In general: events  $A_1, \dots, A_n$  are mutually independent if

$$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

for any subset  $S \subseteq \{1, \dots, n\}$ .

## Random Variables

# Assignment Project Exam Help

- ▶ A random variable  $X$  maps outcomes to real values.
- ▶  $X$  takes on values in  $\text{Val}(X) \subseteq \mathbb{R}$ .

- ▶  $X = k$  is the event that random variable  $X$  takes on value  $k$ .

Discrete RVs:

<https://powcoder.com>

- ▶  $\text{Val}(X)$  is a set

- ▶  $P(X = k)$  can be nonzero

Continuous RVs:

Add WeChat powcoder

- ▶  $\text{Val}(X)$  is a range

- ▶  $P(X = k) = 0$  for all  $k$ .  $P(a \leq X \leq b)$  can be nonzero.

## Probability Mass Function (PMF)

# Assignment Project Exam Help

Given a **discrete** RV  $X$ , a PMF maps values of  $X$  to probabilities.

<https://powcoder.com>

For a valid PMF,  $\sum_{x \in Val(x)} p_X(x) = 1$ .

Add WeChat powcoder

## Cumulative Distribution Function (CDF)

A CDF maps a continuous RV to a probability (i.e.  $\mathbb{R} \rightarrow [0, 1]$ )

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

A CDF must fulfill the following:

- ▶  $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- ▶  $\lim_{x \rightarrow \infty} F_X(x) = 1$
- ▶ If  $a \leq b$ , then  $F_X(a) \leq F_X(b)$  (i.e. CDF must be nondecreasing)

Also note:  $P(a \leq X \leq b) = F_X(b) - F_X(a)$ .

## Probability Density Function (PDF)

PDF of a continuous RV is simply the derivative of the CDF.

# Assignment Project Exam Help

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

Thus,

<https://powcoder.com>

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$$

A valid PDF must be such that

- ▶ for all real numbers  $x$ ,  $f_X(x) \geq 0$ .
- ▶  $\int_{-\infty}^{\infty} f_X(x) dx = 1$

## Expectation

Let  $g$  be an arbitrary real-valued function.

# Assignment Project Exam Help

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

<https://powcoder.com>

- If  $X$  is a continuous RV with PDF  $f_X$ :

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

**Intuitively**, expectation is a weighted average of the values of  $g(x)$ , weighted by the probability of  $x$ .

## Properties of Expectation

For any constant  $a \in \mathbb{R}$  and arbitrary real function  $f$ :

- ▶  $\mathbb{E}[a] = a$

- ▶  $\mathbb{E}[af(X)] = a\mathbb{E}[f(X)]$

# Assignment Project Exam Help

## Linearity of Expectation

Given  $n$  real-valued functions  $f_1(X), \dots, f_n(X)$ ,

$$\mathbb{E}\left[\sum_{i=1}^n f_i(X)\right] = \sum_{i=1}^n \mathbb{E}[f_i(X)]$$

## Law of Total Expectation

Given two RVs  $X, Y$ :

Add WeChat powcoder

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

**N.B.**  $\mathbb{E}[X | Y] = \sum_{x \in Val(x)} x p_{X|Y}(x|y)$  is a function of  $Y$ .  
See Appendix for details :)

## Example of Law of Total Expectation

El Goog sources two batteries,  $A$  and  $B$ , for its phone. A phone with battery  $A$  runs on average 12 hours on a single charge, but only 8 hours on average with battery  $B$ . El Goog puts battery  $A$  in 80% of its phones and battery  $B$  in the rest. If you buy a phone from El Goog, how many hours do you expect it to run on a single charge?

**Solution:** Let  $L$  be the time your phone runs on a single charge. We know the following:

- ▶  $p_X(A) = 0.8$ ,  $p_X(B) = 0.2$ ,
- ▶  $\mathbb{E}[L | A] = 12$ ,  $\mathbb{E}[L | B] = 8$ .

Then, by Law of Total Expectation,

$$\begin{aligned}\mathbb{E}[L] &= \mathbb{E}[\mathbb{E}[L | X]] = \sum_{X \in \{A, B\}} \mathbb{E}[L | X] p_X(X) \\ &= \mathbb{E}[L | A] p_X(A) + \mathbb{E}[L | B] p_X(B) \\ &= 12 \times 0.8 + 8 \times 0.2 = \boxed{11.2}\end{aligned}$$

## Variance

# Assignment Project Exam Help

The variance of a RV  $X$  measures how concentrated the distribution of  $X$  is around its mean.

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

<https://powcoder.com>

**Interpretation:**  $\text{Var}(X)$  is the expected deviation of  $X$  from  $\mathbb{E}[X]$ .

**Properties:** For any constant  $a \in \mathbb{R}$ , real-valued function  $f(X)$

- ▶  $\text{Var}[a] = 0$
- ▶  $\text{Var}[af(X)] = a^2 \text{Var}[f(X)]$

Add WeChat powcoder

## Example Distributions

# Assignment Project Exam Help

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

Read review handout or Sheldon Ross for details <sup>2</sup>

<sup>2</sup>Table reproduced from Maleki & Do's review handout by Koochak & Irvin

## Joint and Marginal Distributions

- ▶ **Joint PMF** for discrete RV's  $X, Y$ :

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

Note that  $\sum_{x \in \text{Val}(X)} \sum_{y \in \text{Val}(Y)} p_{XY}(x, y) = 1$

- ▶ **Marginal PMF** of  $X$ , given joint PMF of  $X, Y$ :

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

<https://powcoder.com>

- ▶ **Joint PDF** for continuous  $X, Y$ :

Add WeChat  powcoder

Note that  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$

- ▶ **Marginal PDF** of  $X$ , given joint PDF of  $X, Y$ :

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

## Joint and Marginal Distributions for Multiple RVs

- ▶ **Joint PMF** for discrete RV's  $X_1, \dots, X_n$ :

$$p(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

**Assignment Project Exam Help**

- ▶ **Marginal PMF** of  $X_1$ , given joint PMF of  $X_1, \dots, X_n$ :

$$p_{X_1}(x_1) = \sum_{x_2} \dots \sum_{x_n} p(x_1, \dots, x_n)$$

<https://powcoder.com>

- ▶ **Joint PDF** for continuous RV's  $X_1, \dots, X_n$ :

Add WeChat  powcoder

Note that  $\int_{x_1} \int_{x_2} \dots \int_{x_n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$

- ▶ **Marginal PDF** of  $X_1$ , given joint PDF of  $X_1, \dots, X_n$ :

$$f_{X_1}(x_1) = \int_{x_2} \dots \int_{x_n} f(x_1, \dots, x_n) dx_2 \dots dx_n$$

## Expectation for multiple random variables

Given two RV's  $X, Y$  and a function  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  of  $X, Y$ ,

- ▶ for discrete  $X, Y$ :

Assignment Project Exam Help

$$\mathbb{E}[g(X, Y)] := \sum_{x \in Val(x)} \sum_{y \in Val(y)} g(x, y) p_{XY}(x, y)$$

- ▶ for continuous  $X, Y$ :

$$\mathbb{E}[g(X, Y)] := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

Add WeChat powcoder

These definitions can be extended to multiple random variables in the same way as in the previous slide. For example, for  $n$  continuous RV's  $X_1, \dots, X_n$  and function  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ :

$$\mathbb{E}[g(X)] = \int \int \dots \int g(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1, \dots, dx_n$$

## Covariance

**Assignment Project Exam Help**

**Intuitively:** measures how much one RV's value tends to move with another RV's value. For RV's  $X, Y$ :

$$\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

- ▶ If  $\text{Cov}[X, Y] < 0$ , then  $X$  and  $Y$  are negatively correlated
- ▶ If  $\text{Cov}[X, Y] > 0$ , then  $X$  and  $Y$  are positively correlated
- ▶ If  $\text{Cov}[X, Y] = 0$ , then  $X$  and  $Y$  are uncorrelated

**Add WeChat powcoder**

## Properties Involving Covariance

- If  $X \perp Y$ , then  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ . Thus,

Assignment Project Exam Help

This is unidirectional!  $\text{Cov}[X, Y] = 0$  **does not imply**  $X \perp Y$

- Variance of two variables:  
 $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$

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

i.e. if  $X \perp Y$   $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

- Special Case:

$$\text{Cov}[X, X] = \mathbb{E}[XX] - \mathbb{E}[X]\mathbb{E}[X] = \text{Var}[X]$$

## Conditional distributions for RVs

Works the same way with *RV's* as with events:

- ▶ For discrete  $X, Y$ :

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

- ▶ For continuous  $X, Y$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

- ▶ Add WeChat powcoder

In general, for continuous  $X_1, \dots, X_n$ .

$$f_{X_1|X_2, \dots, X_n}(x_1|x_2, \dots, x_n) = \frac{f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)}{f_{X_2, \dots, X_n}(x_2, \dots, x_n)}$$

## Bayes' Rule for RVs

# Assignment Project Exam Help

Also works the same way for RV's as with events:

- ▶ For discrete  $X, Y$ :

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y)p_Y(y)}{\sum_{y' \in V_a(Y)} p_{X|Y}(x|y')p_Y(y')}$$

<https://powcoder.com>

- ▶ For continuous  $X, Y$ :

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y')f_Y(y')dy'}$$

Add WeChat **powcoder**

# Assignment Project Exam Help

Also works the same way as with events:

$$\begin{aligned}f(x_1, x_2, \dots, x_n) &\stackrel{def}{=} f(x_1)f(x_2|x_1)\dots f(x_n|x_1, x_2, \dots, x_{n-1}) \\&= f(x_1) \prod_{i=2}^n f(x_i|x_1, \dots, x_{i-1})\end{aligned}$$

Add WeChat powcoder

## Independence for RVs

- ▶ For  $X \perp Y$  to hold, it must be that  $F_{X|Y}(x, y) = F_X(x)F_Y(y)$   
FOR ALL VALUES of  $x, y$

# Assignment Project Exam Help

- ▶ Since  $f_{Y|X}(y|x) = f_Y(y)$  if  $X \perp Y$ , chain rule for mutually independent  $X_1, \dots, X_n$  is:

$$f(x_1, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n) = \prod_{i=1}^n f(x_i)$$

(very important assumption for a Naive Bayes classifier!)

## Random Vectors

Given  $n$  RV's  $X_1, \dots, X_n$ , we can define a random vector  $X$  s.t.

Assignment Project Exam Help

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Note: all the notions of joint PDF/CDF will apply to  $X$

Given  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , we have:

Add WeChat powcoder

$$g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_m(x) \end{bmatrix}, \mathbb{E}[g(X)] = \begin{bmatrix} \mathbb{E}[g_1(X)] \\ \mathbb{E}[g_2(X)] \\ \vdots \\ \mathbb{E}[g_m(X)] \end{bmatrix}.$$

## Covariance Matrices

For a random vector  $X \in \mathbb{R}^n$ , we define its **covariance matrix**  $\Sigma$  as the  $n \times n$  matrix whose  $ij$ -th entry contains the covariance between  $X_i$  and  $X_j$ .

**Assignment Project Exam Help**

$$\Sigma = \begin{bmatrix} \text{Cov}[X_1, X_1] & \dots & \text{Cov}[X_1, X_n] \\ \vdots & \ddots & \vdots \\ \text{Cov}[X_n, X_1] & \dots & \text{Cov}[X_n, X_n] \end{bmatrix}$$

<https://powcoder.com>

applying linearity of expectation and the fact that  $\text{Cov}[X_i, X_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$ , we obtain

**Add WeChat powcoder**

$$\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$$

### Properties:

- ▶  $\Sigma$  is symmetric and PSD
- ▶ If  $X_i \perp X_j$  for all  $i, j$ , then  $\Sigma = \text{diag}(\text{Var}[X_1], \dots, \text{Var}[X_n])$

## Multivariate Gaussian

The multivariate Gaussian  $X \sim \mathcal{N}(\mu, \Sigma)$ ,  $X \in \mathbb{R}^n$ :

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

Assignment Project Exam Help

<https://powcoder.com>

The univariate Gaussian  $X \sim \mathcal{N}(\mu, \sigma^2)$ ,  $X \in \mathbb{R}$  is just the special case of the multivariate Gaussian when  $n = 1$ .

Add WeChat powcoder

Notice that if  $\Sigma \in \mathbb{R}^{1 \times 1}$ , then  $\Sigma = \text{Var}[X_1] = \sigma^2$ , and so

- ▶  $\Sigma^{-1} = \frac{1}{\sigma^2}$
- ▶  $\det(\Sigma)^{\frac{1}{2}} = \sigma$

## Some Nice Properties of MV Gaussians

- ▶ Marginals and conditionals of a joint Gaussian are Gaussian
- ▶ A  $d$ -dimensional Gaussian  $X \in \mathcal{N}(\mu, \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2))$  is equivalent to a collection of  $d$  **independent** Gaussians  $X_i \in \mathcal{N}(\mu_i, \sigma_i^2)$ . This results in isocontours aligned with the coordinate axes.
- ▶ In general, the isocontours of a MV Gaussian are  $n$ -dimensional ellipsoids with principal axes in the directions of the eigenvectors of covariance matrix  $\Sigma$  (remember,  $\Sigma$  is PSD, so all  $n$  eigenvectors are non-negative). The axes' relative lengths depend on the eigenvalues of  $\Sigma$ .

# Visualizations of MV Gaussians

Effect of changing variance

Assignment Project Exam Help

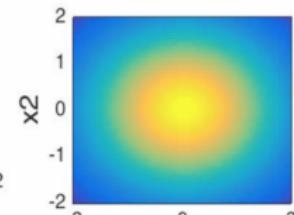
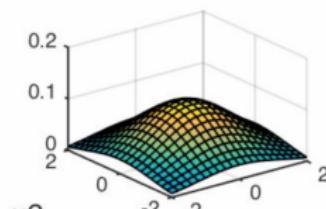
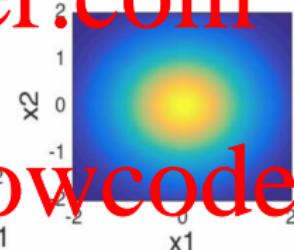
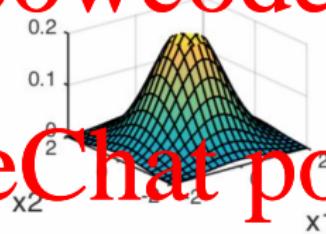
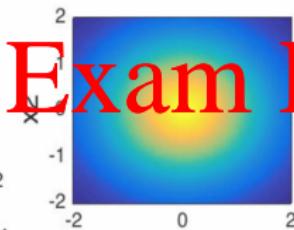
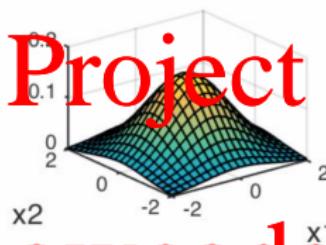
$$\mu = [0 \ 0]^T$$

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

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

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

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



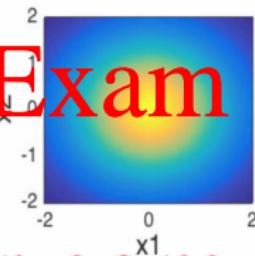
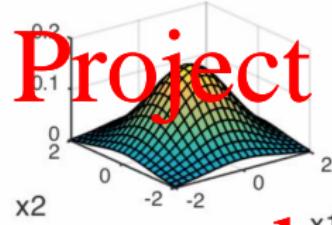
<https://powcoder.com>  
Add WeChat powcoder

## Visualizations of MV Gaussians

If  $\text{Var}[X_1] \neq \text{Var}[X_2]$ :

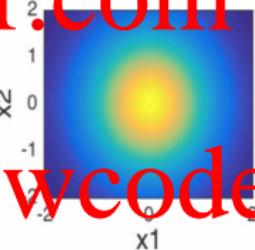
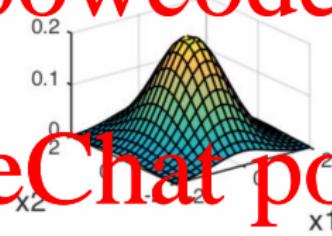
$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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



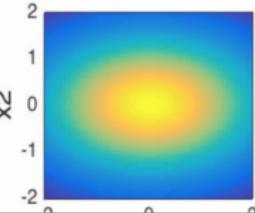
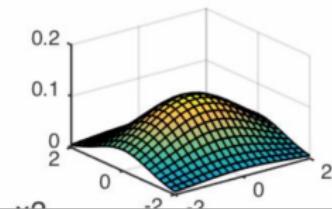
$$\Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix}$$

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



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

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



## Visualizations of MV Gaussians

If  $X_1$  and  $X_2$  are positively correlated:

Assignment Project Exam Help

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

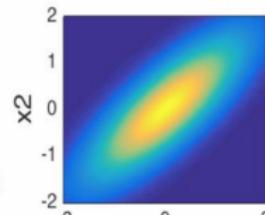
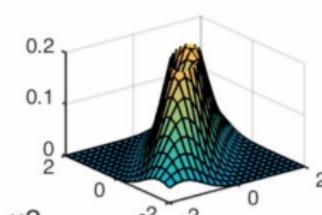
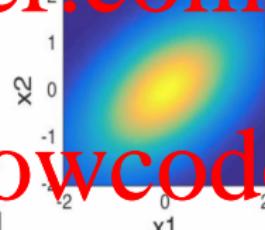
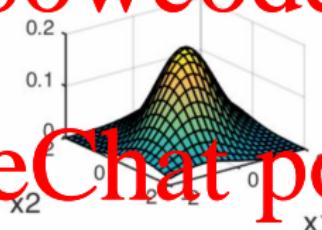
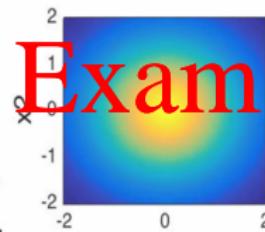
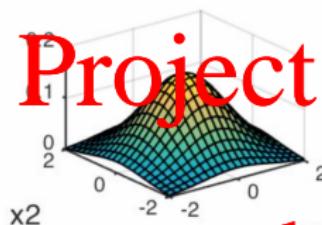
$$\mu = [0 \ 0]^T$$

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

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

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

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



## Visualizations of MV Gaussians

If  $X_1$  and  $X_2$  are negatively correlated:

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

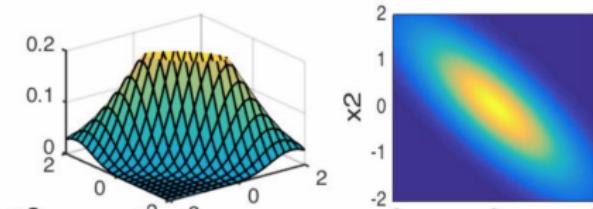
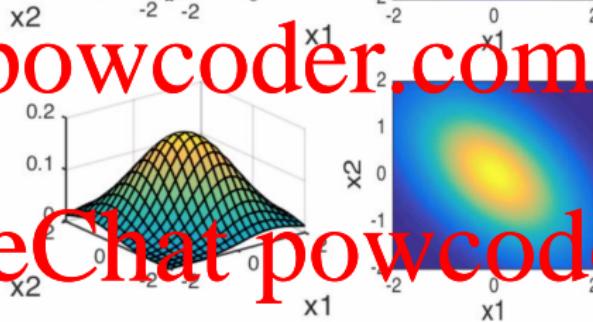
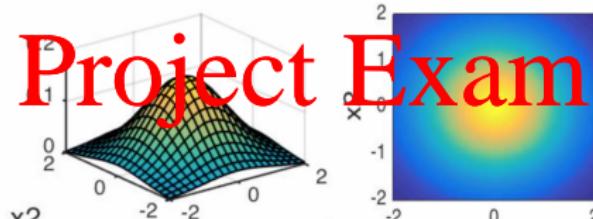
$$\mu = [0 \ 0]^T$$

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

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

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

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



Thank you and good luck!

# Assignment Project Exam Help

For further reference, consult the following CS229 handouts

- ▶ Probability Theory Review
- ▶ The MV Gaussian Distribution
- ▶ More on Gaussian Distribution

For a comprehensive treatment, see

- ▶ Sheldon Ross, A First Course in Probability

Add WeChat powcoder

## Appendix: More on Total Expectation

Why is  $\mathbb{E}[X|Y]$  a function of  $Y$ ? Consider the following:

- ▶  $\mathbb{E}[X|Y = y]$  is a scalar that only depends on  $y$ .
- ▶ Thus,  $\mathbb{E}[X|Y]$  is a random variable that only depends on  $Y$ . Specifically,  $\mathbb{E}[X|Y]$  is a function of  $Y$  mapping  $Val(Y)$  to the real numbers.

An example: Consider RV  $X$  such that

$$X = Y^2 + \epsilon$$

such that  $\epsilon \sim \mathcal{N}(0, 1)$  is a standard Gaussian. Then,

- ▶  $\mathbb{E}[X|Y] = Y^2$
- ▶  $\mathbb{E}[X|Y = y] = y^2$

## Appendix: More on Total Expectation

A derivation of Law of Total Expectation for discrete  $X, Y$ :<sup>3</sup>

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}\left[\sum_x xP(X=x|Y)\right] \quad (1)$$

$$= \sum_y \sum_x xP(X=x|Y)P(Y=y) \quad (2)$$

$$= \sum_y \sum_x xP(X=x, Y=y) \quad (3)$$

$$= \sum_x x \sum_y P(X=x, Y=y) \quad (4)$$

$$= \sum_x xP(X=x) - \boxed{\mathbb{E}[X]} \quad (5)$$

where (1), (2), and (5) result from the definition of expectation, (3) results from the definition of cond. prob., and (5) results from marginalizing out  $Y$ .

---

<sup>3</sup>from slides by Koochak & Irvin

## Appendix: A proof of Conditioned Bayes Rule

Repeatedly applying the definition of conditional probability, we have:<sup>4</sup>

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

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

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

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

$$= P(a|b,c)$$

Add WeChat powcoder

---

<sup>4</sup>from slides by Koochak & Irvin