Continuous Inference
Chris Piech
CS109, Stanford University
<Review>
PDFs of Multiple Continuous Variables

If two random variables $X$ and $Y$ are jointly continuous, then there exists a joint probability density function $f_{X,Y}$ defined over $-\infty < x, y < \infty$ such that:

$$P(a_1 \leq X \leq a_2, \ b_1 \leq Y \leq b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y) dy \ dx$$
CDFs of Multiple Continuous Variables

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \\
F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1) \]
Independence of Multiple Continuous Variables

Two continuous random variables $X$ and $Y$ are independent if:

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y) \quad \forall x, y$$

Equivalently:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x, y$$

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall x, y$$
Bivariate Gaussians

1. PDF

2. PDF

3. PDF

4. PDF
Bivariate Gaussians

Two continuous variables \(X_1, X_2 \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)

Sometimes written equivalently as:

\[
\begin{align*}
\vec{X} & \sim N_k(\vec{\mu}, \Sigma) \\
\Sigma &= \begin{bmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}
\end{align*}
\]
Bivariate Gaussians PDF

\( X_1 \) and \( X_2 \) follow a bivariate normal distribution if their joint PDF \( f \) is

\[
f(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left( \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)}
\]
But what if $\rho = 0$?

$X_1$ and $X_2$ follow a bivariate normal distribution if their joint PDF $f$ is

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

$$f(x_1, x_2) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \quad \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x_2-\mu_2)^2}{2\sigma_2^2}}$$
Independent Multivariate Gaussians

$X_1$ and $X_2$ are independent with marginal distributions. Or equivalently if $\rho = 0$, $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

Joint PDF

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

Joint CDF

$$F(x_1, x_2) = \Phi\left(\frac{x_1 - \mu_1}{\sigma_1}\right) \cdot \Phi\left(\frac{x_2 - \mu_2}{\sigma_2}\right)$$
Multivariate Normal Distributions are Rad 😊

The Data

A 2D scatter plot with three "clusters"

The Joint Model

A mixture of 3 bi-variate Normal distributions

See you in CS221, Gaussian Mixture Models
In image processing, a Gaussian blur is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise.
Gaussian blur

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that $X$ and $Y$ are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter $\sigma$

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) - F_{X,Y}(a_2,b_1) + F_{X,Y}(a_1,b_1)$$

Weight matrix:

Center pixel: (0, 0)
Pixel bounds: 
$$-0.5 < x \leq 0.5$$
$$-0.5 < y \leq 0.5$$

- Independently: $X \sim \mathcal{N}(0, 3^2), Y \sim \mathcal{N}(0, 3^2)$
- Jointly: $X, Y \sim \mathcal{N}_2(\mu_x = 0, \mu_y = 0, \sigma_x^2 = 3^2, \sigma_y^2 = 3^2, \rho = 0)$

What is the weight of the center pixel?
Gaussian blur

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that \(X\) and \(Y\) are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter \(\sigma\)

\[f_{X,Y}(x, y) = \frac{1}{2\pi \cdot 3^2} e^{-(x^2+y^2)/2\cdot 3^2}\]

Weight matrix:

- Center pixel: \((0, 0)\)
- Pixel bounds:
  \(-0.5 < x \leq 0.5\)
  \(-0.5 < y \leq 0.5\)

\[F_{X,Y}(a_1, b_1, a_2, b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1)\]
Gaussian blur

In a Gaussian blur, for every pixel:

- Weight each pixel by the probability that $X$ and $Y$ are both within the pixel bounds
- The weighting function is a Bivariate Gaussian (Normal) standard deviation parameter $\sigma$

\[ P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) - F_{X,Y}(a_2, b_1) + F_{X,Y}(a_1, b_1) \]

Weight matrix:
Center pixel: (0, 0)
Pixel bounds:
\[-0.5 < x \leq 0.5 \]
\[-0.5 < y \leq 0.5 \]

\[ \rightarrow \text{Independently: } X \sim \mathcal{N}(0, 3^2), Y \sim \mathcal{N}(0, 3^2) \]
\[ \rightarrow \text{Jointly: } X, Y \sim \mathcal{N}_2(\mu_x = 0, \mu_y = 0, \sigma^2_x = 3^2, \sigma^2_y = 3^2, \rho = 0) \]

What is the weight of the center pixel?

\[ P(-0.5 < X \leq 0.5, -0.5 < Y \leq 0.5) = 0.206 \]
Is there a version of Bayes' Theorem for continuous Random Variables?
Conditional probability and Bayes’ Theorem

Definition

\[ P(F|E) = \frac{P(E \cap F)}{P(E)} \]

Prior: some prob. of event \( F \)

Independence

\[ P(F|E) = P(F) \]

Sample space doesn’t need to be scaled

Bayes’ Theorem

\[ P(F|E) = \frac{P(F)P(E|F)}{P(E)} \]

Posterior: prob. of \( F \) knowing that \( E \) happened

Likelihood

Scaling to the correct sample space
Multiple Bayes’ Theorems

with events

\[
P(F|E) = \frac{P(F)P(E|F)}{P(E)}
\]

with discrete RVs

\[
p_{Y|X}(y|x) = \frac{p_Y(y)p_{X|Y}(x|y)}{p_X(x)}
\]

with continuous RVs
Relative Probability of Continuous Variables

\( X = \text{time to finish pset 3} \)
\( X \sim N(10, 2) \)

How much more likely are you to complete in 10 hours than in 5?

\[
\frac{P(X = 10)}{P(X = 5)} = \frac{\varepsilon f(X = 10)}{\varepsilon f(X = 5)}
\]

\[
= \frac{f(X = 10)}{f(X = 5)}
\]

\[
= \frac{\frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(10-\mu)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(5-\mu)^2}{2\sigma^2}}}
\]

\[
= \frac{\frac{1}{\sqrt{4\pi}}e^{-\frac{(10-10)^2}{4}}}{\frac{1}{\sqrt{4\pi}}e^{-\frac{(5-10)^2}{4}}}
\]

\[
= \frac{e^0}{e^{-\frac{25}{4}}} = 518
\]
Bayes with Continuous Random Variables

Let $X$ and $Y$ be continuous random variables

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

\[ f_{X|Y}(x | y) \cdot \epsilon_x = \frac{f_{X,Y}(x, y) \cdot \epsilon_x \cdot \epsilon_y}{f_Y(y) \cdot \epsilon_y} \]

\[ f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \]
Multiple Bayes’ Theorems

with events

\[ P(F|E) = \frac{P(F)P(E|F)}{P(E)} \]

with discrete RVs

\[ p_{Y|X}(y|x) = \frac{p_Y(y)p_{X|Y}(x|y)}{p_X(x)} \]

with continuous RVs

\[ f_{Y|X}(y|x) = \frac{f_Y(y)f_{X|Y}(x|y)}{f_X(x)} \]

Really all the same idea!
Mixing Discrete and Continuous

Let $X$ be a continuous random variable

Let $N$ be a discrete random variable

$$\Pr(X = x \mid N = n) = \frac{\Pr(N = n \mid X = x) \Pr(X = x)}{\Pr(N = n)}$$

$$p_{X \mid N}(x \mid n) = \frac{p_{N \mid X}(n \mid x)p_X(x)}{p_N(n)}$$

$$f_{X \mid N}(x \mid n) \cdot \epsilon_x = \frac{p_{N \mid X}(n \mid x)f_X(x) \cdot \epsilon_x}{p_N(n)}$$

$$f_{X \mid N}(x \mid n) = \frac{p_{N \mid X}(n \mid x)f_X(x)}{p_N(n)}$$
All the Bayes Belong to Us!

M, N are discrete. X, Y are continuous

\[
p_{M|N}(m|n) = \frac{P_{N|M}(n|m)p_M(m)}{p_N(n)}
\]

\[
f_{X|N}(x|n) = \frac{P_{N|X}(n|x)f_X(x)}{P_N(n)}
\]

\[
p_{N|X}(n|x) = \frac{f_{X|N}(x|n)p_N(n)}{f_X(x)}
\]

\[
f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}
\]

OG Bayes

Mix Bayes #1

Mix Bayes #2

All Continuous

Chris Piech and Jerry Cain, CS109, 2020
Stanford Acuity Test

1. Take an eye exam on this website
2. Connect your phone
3. Visualize the math
A user is shown a letter at font size 3 and gets it wrong. What is your new belief that their visual ability is 3?

Aside: font size = 3, means it is 3x what someone with healthy vision can see clearly. Visual ability = 3 means the person can see font size 3 with 80% accuracy.
First let's define a few variables.

Random Variables

- \( A \): Ability to see
- \( Y \): Indicates that they answered a question correct

Numbers

- \( x \): Font size shown

Probability Question

\[
f(A = a | Y = 0)
\]
Your prior belief in ability to see is distributed as a gumbel

\[ f(A = a) = 3.3 \cdot e^{-0.3a-0.33} - e^{(-0.3a-0.33)} \]

Curious?

https://en.wikipedia.org/wiki/Gumbel_distribution
What is the probability you are correct, given ability

\[ P(Y = 1|A = a) = \max(0.25, 1 - 0.75 \cdot 0.27^{\frac{x-a}{0.2a}}) \]
Bring it all together.

What we **know**

\[
f(A = a) = 3.3 \cdot e^{-0.3a - 0.33} - e^{(-0.3a - 0.33)}
\]

\[
P(Y = 1|A = a) = \max(0.25, 1 - 0.75 \cdot 0.27^{\frac{x - a}{0.2a}})
\]

What we **want**

\[
f(A = 3|Y = 0) = \frac{[1 - P(Y = 1|A = 3)] \cdot f(A = 3)}{P(Y = 0)}
\]

\[
= \frac{[1 - P(Y = 1|A = 3)] \cdot f(A = 3)}{\int_a P(Y = 0|A = a) \, da}
\]

\[
= K \cdot [1 - P(Y = 1|A = 3)] \cdot f(A = 3)
\]

Font-size was 3
An Updated Belief

A user is shown a letter at font size 3 and gets it wrong. What is your new belief that their visual ability is 3?
So what?

Actual model also included
+ a probability of "slip"
+ an intelligent algorithm for choosing the next letter size
Tracking in 2D Space?
Tracking in 2-D space

• Before measuring, we have some prior belief about the 2-D location of an object, \((X, Y)\).

• We observe some noisy measurement \(D = 4\), the Euclidean distance of the object to a satellite.

• After the measurement, what is our updated (posterior) belief of the 2-D location of the object?
Before measuring, we have some prior belief about the 2-D location of an object, \((X, Y)\).
Tracking in 2-D space

- Before measuring, we have some **prior belief** about the 2-D location of an object, \((X, Y)\).

- We observe some noisy **measurement** \(D = 4\), the Euclidean distance of the object to a satellite.

Let \(D = \) observed distance from the satellite. Observed distance is true distance plus noise. Noise is a **standard normal**.
Tracking in 2-D space

- You have a prior belief about the 2-D location of an object, \((X, Y)\).
- You observe a noisy distance measurement, \(D = 4\).
- What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Recall Bayes terminology:

\[
f_{X,Y|D}(x, y|d) = \frac{f_{D|X,Y}(d|x, y)f_{X,Y}(x, y)}{f_D(d)}
\]

- posterior belief
- likelihood (of evidence)
- prior belief
- normalization constant
1. Define prior

You have a **prior belief** about the 2-D location of an object, \((X, Y)\).

Let \((X, Y) = \text{object's 2-D location.} \)
(your satellite is at \((0,0)\))

Suppose the prior distribution is a symmetric bivariate normal distribution:

\[
f_{X,Y}(x, y) = \frac{1}{2\pi \sigma^2} e^{-\frac{(x-\mu_x)^2 + (y-\mu_y)^2}{2\sigma^2}}
\]

= \(K_1 \cdot e^{-\frac{(x-3)^2 + (y-3)^2}{8}}\)

normalizing constant
2. Define likelihood

You observe a noisy distance measurement, $D = 4$.

If you knew your actual location $(x, y)$, you could say how likely a measurement $D = 4$ is:

Let $D =$ distance from the satellite (radially).

Suppose you knew your actual position: $(x, y)$.

- $D$ is still noisy! Suppose noise is standard normal.
- On average, $D$ is your true Euclidean distance: $\sqrt{x^2 + y^2}$
2. Define likelihood

You observe a noisy distance measurement, \( D = 4 \).

If you knew your actual location \((x, y)\), you could say how likely a measurement \( D = 4 \) is:

Let \( D \) = distance from the satellite (radially).

Suppose you knew your actual position: \((x, y)\).

- \( D \) is still noisy! Suppose noise is **standard normal**.
- On average, \( D \) is your true Euclidean distance: \( \sqrt{x^2 + y^2} \)

\[
D | X, Y \sim N(\mu = (A), \sigma^2 = (B))
\]

\[
f_{D|X,Y}(D = d | X = x, Y = y) = \frac{1}{(C) \sqrt{2\pi}} e^{\{ (D) \}}
\]
Define likelihood

You observe a noisy distance measurement, \( D = 4 \).

If you knew your actual location \((x, y)\), you could say how likely a measurement \( D = 4 \) is:

Let \( D = \text{distance from the satellite (radially)} \).

Suppose you knew your actual position: \((x, y)\).
- \( D \) is still noisy! Suppose noise is \textbf{standard normal}.
- On average, \( D \) is your true Euclidean distance: \( \sqrt{x^2 + y^2} \)

\[
D \mid X, Y \sim N \left( \mu = \sqrt{x^2 + y^2}, \sigma^2 = 1 \right)
\]

\[
f_{D \mid X, Y}(D = d \mid X = x, Y = y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(d-\sqrt{x^2+y^2})^2}{2}} = K_2 \cdot e^{-\frac{(d-\sqrt{x^2+y^2})^2}{2}}
\]

normalizing constant
3. Compute posterior

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Compute:

Posterior belief

\[
f_{X,Y|D}(x, y|4) = f_{X,Y|D}(X = x, Y = y|D = 4)
\]
3. Compute posterior

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

Compute:

Posterior belief

\[ f_{X,Y|D}(x,y|4) = f_{X,Y|D}(X = x, Y = y|D = 4) \]

Know:

Prior belief

\[ f_{X,Y}(x,y) = K_1 \cdot e^{-\frac{(x-3)^2 + (y-3)^2}{8}} \]

Observation likelihood

\[ f_{D|X,Y}(d|x,y) = K_2 \cdot e^{-\frac{(d-x^2+y^2)^2}{2}} \]

Tips

• Use Bayes’ Theorem!
• \( f_D(4) \) is just a scaling constant. Why?
• How can we approximate the final scaling constant with a computer?
Deep breath
Tracking in 2-D space

What is your updated (posterior) belief of the 2-D location of the object after observing the measurement?

\[
f_{X,Y|D}(X = x, Y = y|D = 4) = \frac{f_{D|X,Y}(D = 4|X = x, Y = y)f_{X,Y}(x, y)}{f(D = 4)}
\]

likelihood of \(D = 4\) prior belief

\[
K_2 \cdot e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2}} \cdot K_1 \cdot e^{-\frac{[(x-3)^2+(y-3)^2]}{8}}
\]

\[
K_3 \cdot e^{-\left[\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{[(x-3)^2+(y-3)^2]}{8}\right]}\]

For your notes...

Key: Once we know the part dependent on \(x, y\), we can computationally approximate \(K_4\) such that \(f_{X,Y|D}\) is a valid PDF.
Tracking in 2-D space

With this continuous version of Bayes’ theorem, we can explore new domains.

• Before measuring, we have some **prior belief** about the 2-D location of an object, \((X, Y)\).

• We observe some noisy **measurement** of the distance of the object to a satellite.

• After the measurement, what is our **updated (posterior) belief** of the 2-D location of the object?
Tracking in 2-D space: Posterior belief

Prior belief

Top-down view

3-D view

Posterior belief

Top-down view

3-D view

\[ f_{X,Y}(x, y) = K_1 \cdot e^{-\frac{(x-3)^2+(y-3)^2}{8}} \]

\[ f_{X,Y|D}(x, y|4) = K_4 \cdot e^{-\left(\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}\right)} \]
How’d you compute that $K_4$?

To be a valid conditional PDF,\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y|D}(x, y|4) \, dx \, dy = 1
\]

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_4 \cdot e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}} \, dx \, dy = 1
\]

\[
\frac{1}{K_4} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}} \, dx \, dy
\]

(pull out $K_4$, divide)

Approximate:

\[
\frac{1}{K_4} \approx \sum_{x} \sum_{y} e^{-\frac{(4-\sqrt{x^2+y^2})^2}{2} + \frac{(x-3)^2+(y-3)^2}{8}} \Delta x \Delta y
\]

Use a computer!
Tracking in 2D Space