Math Camp

Justin Grimmer

Professor
Department of Political Science
Stanford University

September 18th, 2018

Questions?

1) What is a continuous random variable?
2) What does it mean when we say $X \sim \text{Normal}(\mu, \sigma^2)$
3) Explain why the pdf and cdf contain the same information
4) Explain why the height of the pdf isn't a probability
5) Suppose $Z \sim \text{Normal}(0, 1)$. What is $Y = aZ + b$?

Where We've Been, Where We're Going

Multivariate Distributions

1) Joint Density
2) Covariance, Marginalization
3) Independence of Random Variables
4) Properties of Sums of Random Variables
5) The Multivariate Normal Distribution and You

Continuous Random Variable

Definition

X is a continuous random variable if there exists a nonnegative function defined for all $x \in \Re$ having the property for any (measurable) set of real numbers B,

$$P(X \in B) \quad = \quad \int_B f(x) dx$$

We'll call $f(\cdot)$ the probability density function for X.

Definition

Multivariate Distribution *We will say that X and Y are jointly continuous if, for all $x \in \Re$ and $y \in \Re$, there exists a function $f(x, y)$ such ever set $C \subset \Re^2$,*

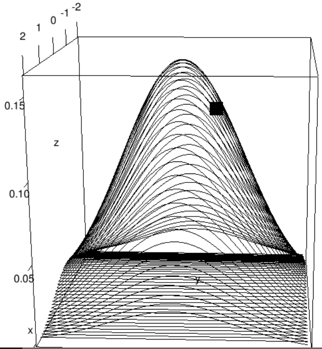
$$P\{(X, Y) \in C\} \ = \ \int_B \int_A f(x, y) dxdy$$

What is $C \subset R^2$?

- $R^2 = R \underbrace{\times}_{\text{Cartesian Product}} R$
    - This is the 2-d plane (your piece of paper)
- C is a subset of the 2-d plane
    - $C = \{x, y : x \in [0, 1], y \in [0, 1]\}$
    - $C = \{x, y : x^2 + y^2 \leq 1\}$
    - $C = \{x, y : x > y, x, y \in (0, 2)\}$
    - $C = \{x, y : x \in A, y \in B\}$

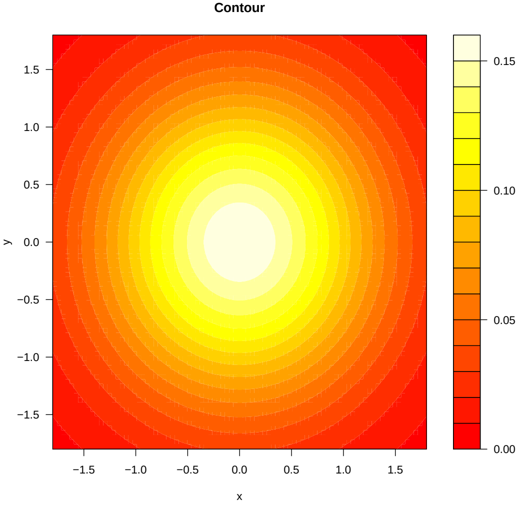
Examples of Joint PDFs

- We're going to focus (initially) on pdfs of two random variables
- Consider a function $f : \Re \times \Re \rightarrow \Re$
  - Input: an x value and a y value.
  - Output: a number from the real line
  - f(x,y) = a



Equivalently: Contour Plots

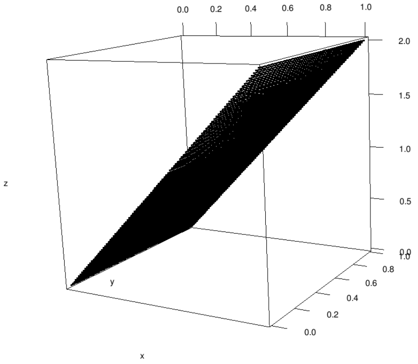
Aerial view of probability density function: contour plots



Example 3D-Contour Plots

Joint distribution of X and Y.

3) $f(x, y) = x + y$, if $x \in [0, 1], y \in [0, 1]$



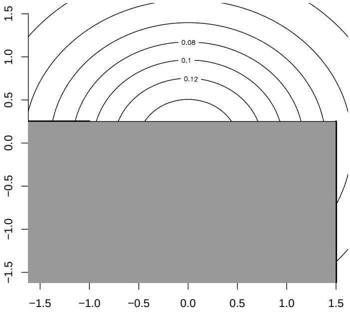
Definition

Multivariate Cumulative Density Function

For jointly continuous random variables X and Y define, $F(b,a)$ as

$$F(b,a) = P\{X \leq b, Y \leq a\} \;=\; \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dxdy$$

A Picture



Examples:

- $F(1.5, 0.25)$

$$F(b,a) \ = \ \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dxdy$$

Marginalization

Definition

Moving from Joint Distributions to Univariate PDFs

Define $f_X(x)$ *as the pdf for* X,

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

Similarly, define $f_Y(y)$ *as the pdf for* Y,

$$f_Y(y) \quad = \quad \int_{-\infty}^{\infty} f(x, y) dx$$

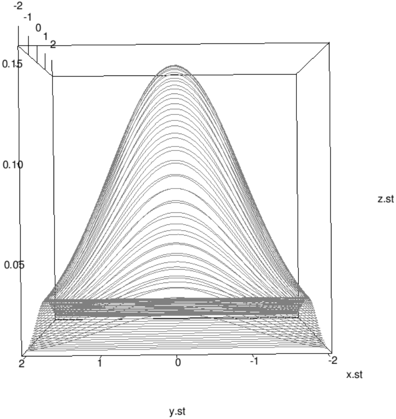
Conditional Probability Distribution Function

Definition

Suppose X and Y are continuous random variables with joint pdf $f(x, y)$. Then define the conditional probability function $f(x|y)$ as

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

A Picture



Why Does Marginalization Work?

Begin with discrete case.

Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Suppose that the distribution allocates its mass at $x_1, x_2, \ldots, x_M$ and $y_1, y_2, \ldots, y_N$.

Define the conditional mass function $P(X = x|Y = y)$ as,

$$\begin{aligned} P(X = x|Y = y) \equiv &= p(x|y) \\ &= p(x, y)/p(y) \end{aligned}$$

Then it follows that:

$$p(x, y) = p(x|y)p(y)$$

Marginalizing over y to get $p(x)$ is then,

$$p(x_j) = \sum_{i=1}^{N} p(x_j|y_i)p(y_i)$$

A Table

	Y = 0	Y = 1	
X = 0	0.01	0.05	?
X = 1	0.25	0.69	?
	0.26	0.74	

$$\begin{aligned} p_X(0) &= p(0|y=0)p(y=0) + p(0|y=1)p(y=1) \\ &= \frac{0.01}{0.26} \times 0.26 + \frac{0.05}{0.74} \times 0.74 \\ &= 0.06 \end{aligned}$$

$$\begin{aligned} p_X(1) &= p(1|y=0)p(y=0) + p(1|y=1)p(y=1) \\ &= \frac{0.25}{0.26} \times 0.26 + \frac{0.69}{0.74} \times 0.74 \\ &= 0.94 \end{aligned}$$

Move to the Continuous Case

For jointly distributed continuous random variables X and Y define,

$$f_{X|Y}(x|y) \quad = \quad \frac{f(x,y)}{f_Y(y)}$$

Then, analogously, we can define

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

In words:

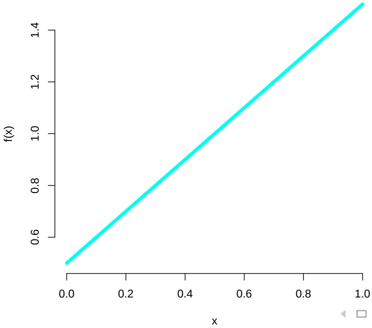
- Think of $f_{X|Y}(x|y)$ as the pdf for X at a value of Y.
- We average over those pdfs to get the final pdf for X (want densities where there is lots of area of Y to receive lots of weight, the densities without much area from Y should receive little weight)

A (Simple) Example

Suppose X and Y are jointly continuous and that

$$
\begin{aligned}
f(x,y) &= x + y \text{ , if } x \in [0,1], y \in [0,1] \\
&= 0 \text{ , otherwise}
\end{aligned}
$$

We want $f_X(x)$. Assume we have $f_Y(y) = 1/2 + y$.
Then: $f(x|y) = \frac{x+y}{1/2+y}$. $f(x) = \int_0^1 f(x|y)f(y)dy = 1/2 + x$



A (Simple) Example

Example

(Ross, Example 1)

Suppose X and Y are jointly distributed with pdf (for all$x > 0, y > 0$)

$$f(x,y) \quad = \quad 2\exp(-x)\exp(-2y)$$

1) Verify this is a pdf

$$\begin{aligned}
\int_0^{\infty}\int_0^{\infty} f(x,y)dxdy &= 2\int_0^{\infty}\int_0^{\infty}\exp(-x)\exp(-2y)dxdy \\
&= 2\int_0^{\infty}\exp(-2y)dy\int_0^{\infty}\exp(-x)dx \\
&= 2(-\frac{1}{2}\exp(-2y)|_0^{\infty})(-\exp(-x)|_0^{\infty}) \\
&= 2\left[(-\frac{1}{2}(\lim_{y\to\infty}\exp(-2y) - 1))(-(\lim_{x\to\infty}\exp(-x) - 1))\right] \\
&= 2\left[-\frac{1}{2}(-1)\times -1(-1)\right] \\
&= 1
\end{aligned}$$

2) Calculate CDF

$$
\begin{aligned}
F(x,y) \equiv P\{X \leq b, Y \leq a\} &= 2\int_0^a \int_0^b \exp(-x)\exp(-2y)dxdy \\
&= 2(\int_0^a \exp(-2y)dy)(\int_0^b \exp(-x)dx) \\
&= 2\left[-\frac{1}{2}(\exp(-2a)-1)\right][-(\exp(-b)-1)] \\
&= [1-\exp(-2a)][1-\exp(-b)]
\end{aligned}
$$

3) Calculate $f_X(x)$ and $f_Y(y)$

$$\begin{aligned}
f_X(x) &= \int_0^{\infty} 2\exp(-x)\exp(-2y)dy \\
&= 2\exp(-x)\int_0^{\infty} \exp(-2y)dy \\
&= 2\exp(-x)\left[-\frac{1}{2}(0-1)\right] \\
&= \exp(-x)
\end{aligned}$$

$$\begin{aligned}
f_Y(y) &= \int_0^{\infty} 2\exp(-x)\exp(-2y)dx \\
&= 2\exp(-2y)\int_0^{\infty} \exp(-x)dx \\
&= 2\exp(-2y)\left[-(0-1)\right] \\
&= 2\exp(-2y)
\end{aligned}$$

Definition

Two random variables X and Y are independent if for any two sets of real numbers A and B,

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

Equivalently we will say X and Y are independent if,

$$f(x, y) = f_X(x)f_Y(y)$$

If X and Y are not independent, we will say they are dependent

Conditional Distribution

If X and Y are independent, then

$$
\begin{aligned}
f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\
&= \frac{f_X(x)f_Y(y)}{f_Y(y)} \\
&= f_X(x)
\end{aligned}
$$

In words: the distribution of X does not change as levels of Y change.

A (Simple) Example of Dependence

Suppose X and Y are jointly continuous and that

$$\begin{aligned} f(x,y) &= x + y \text{ , if } x \in [0,1], y \in [0,1] \\ &= 0 \text{ , otherwise} \end{aligned}$$

$$\begin{aligned} f(x,y) &= x + y \\ f_X(x)f_Y(y) &= (\frac{1}{2} + x)(\frac{1}{2} + y) \\ &= \frac{1}{4} + \frac{x+y}{2} + xy \end{aligned}$$

Intuition: at different levels of X the distribution on Y behaves differently.
X provides information about Y

Expectation

Definition

For jointly continuous random variables X and Y define,

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy \\
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy \\
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy
\end{aligned}
$$

Proposition

Suppose $g : \Re^2 \rightarrow \Re$ (that isn't crazy). Then,

$$
E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy
$$

Covariance

Definition

For jointly continous random variables X *and* Y *define, the covariance of* X *and* Y *as,*

$$\begin{aligned} cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E\left[XY - E[X]Y - E[Y]X + E[X]E[Y]\right] \\ &= E[XY] - 2E[X]E[Y] + E[E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Define the correlation of X *and* Y *as,*

$$cor(X, Y) = \frac{cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Some Observations

Variance is the covariance of a random variable with itself.

$$\begin{aligned} cov(X, X) &= E[XX] - E[X]E[X] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Correlation measures the linear relationship between two random variables

Suppose $X = Y$

$$\begin{aligned} cor(X, Y) &= \frac{cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{Var(X)}{Var(X)} \\ &= 1 \end{aligned}$$

Suppose $X = -Y$

$$\begin{aligned} cor(X, Y) &= \frac{cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{-Var(X)}{Var(X)} \end{aligned}$$

Correlation is Between -1 and 1

$|cor(X, Y)| \leq 1$

- Proof 1: Variance trick
- Proof 2: Cauchy-Schwartz Inequality
    - "Inner product" of any two vectors X and Y is less than or equal to the length of vector X times the length of vector Y

Example: $X + Y$

Suppose X and Y have pdf $x + y$ for $x, y \in [0, 1]$.
$\text{Cov}(X, Y)$

$$
\begin{aligned}
E[XY] &= \int_0^1 \int_0^1 xy(x+y)dxdy \\
&= \int_0^1 \int_0^1 (x^2 y + y^2 x)dxdy \\
&= \int_0^1 (\frac{y}{3} + \frac{y^2}{2})dy \\
&= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \int_0^1 \int_0^1 y(x+y)dxdy \\
&= \frac{7}{12}
\end{aligned}
$$

Example: $X + Y$

$$
\begin{aligned}
\text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
&= \frac{1}{3} - \frac{49}{144} = -\frac{1}{144}
\end{aligned}
$$

$$
\begin{aligned}
\text{Cor}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\
&= \frac{-\frac{1}{144}}{\frac{11}{144}} \\
&= \frac{-1}{11}
\end{aligned}
$$

Sums of Random Variables

Suppose we have a sequence of random variables X_i , $i = 1, 2, \ldots, N$. Suppose that they have joint pdf,

$$f(\boldsymbol{x}) \quad = \quad f(x_1, x_2, \ldots, x_n)$$

1) $E[\sum_{i=1}^N X_i] = \sum_{i=1}^N E[X_i]$
2) $\text{var}(\sum_{i=1}^N X_i) = \sum_{i=1}^N \text{var}(X_i) + 2\sum_{i<j} \text{cov}(X_i, X_j)$

Sums of Random Variables

Proposition

Suppose we have a sequence of random variables X_i , $i = 1, 2, \ldots, N$. Suppose that they have joint pdf,

$$f(\boldsymbol{x}) \quad = \quad f(x_1, x_2, \ldots, x_n)$$

Then

$$E[\sum_{i=1}^{N} X_i] \quad = \quad \sum_{i=1}^{N} E[X_i]$$

Proof.

$$
\begin{aligned}
E[\sum_{i=1}^{N} X_i] &= E[X_1 + X_2 + \ldots + X_N] \\
&= \int_{-\infty}^{\infty} \cdots \int\int_{-\infty}^{\infty} (x_1 + x_2 + \ldots + x_N) f(x_1, x_2, \ldots, x_N) dx_1 dx_2 \ldots dx_N \\
&= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 + \ldots + \int_{-\infty}^{\infty} x_N f_{X_N}(x_N) dx_N \\
&= E[X_1] + E[X_2] + \ldots + E[X_N]
\end{aligned}
$$

$\square$

Sums of Random Variable

Proposition

Suppose X_i is a sequence of random variables. Then

$$var(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} var(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

Sums of Random Variable

Proof.

Consider two random variables, X_1 and X_2. Then,

$$
\begin{aligned}
\text{var}(X_1 + X_2) &= E[(X_1 + X_2)^2] - (E[X_1] + E[X_2])^2 \\
&= E[X_1^2] + 2E[X_1X_2] + E[X_2^2] \\
&\quad -(E[X_1])^2 - 2E[X_1]E[X_2] - 2E[X_2]^2 \\
&= \underbrace{E[X_1^2] - (E[X_1])^2}_{\text{var}(X_1)} + \underbrace{E[X_2^2] - E[X_2]^2}_{\text{var}(X_2)} \\
&\quad +2\underbrace{(E[X_1X_2] - E[X_1]E[X_2])}_{\text{cov}(X_1,X_2)} \\
&= \text{var}(X_1) + \text{var}(X_2) + 2\text{cov}(X_1, X_2)
\end{aligned}
$$

□

Definition

Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ *is a vector of random variables. If* $\boldsymbol{X}$ *has pdf*

$$f(\boldsymbol{x}) = (2\pi)^{-N/2} det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

Then we will say $\boldsymbol{X}$ *is a Multivariate Normal Distribution,*

$$\boldsymbol{X} \sim \textit{Multivariate Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- Regularly used for likelihood, Bayesian, and other parametric inferences

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} f(x_1, x_2) &= (2\pi)^{-2/2} 1^{-1/2} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \mathbf{0})' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{x} - \mathbf{0})\right)\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) \end{aligned}$$

$\rightsquigarrow$ product of univariate standard normally distributed random variables

Standard Multivariate Normal

Definition

Suppose $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_N)$ *is*

$$\boldsymbol{Z} \sim \textit{Multivariate Normal}(\mathbf{0}, \boldsymbol{I}_N).$$

Then we will call $\boldsymbol{Z}$ *the standard multivariate normal.*

Properties of the Multivariate Normal Distribution

Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$

$$
\begin{aligned}
E[\boldsymbol{X}] &= \boldsymbol{\mu} \\
\text{cov}(\boldsymbol{X}) &= \boldsymbol{\Sigma}
\end{aligned}
$$

So that,

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
\text{var}(X_1) & \text{cov}(X_1, X_2) & \ldots & \text{cov}(X_1, X_N) \\
\text{cov}(X_2, X_1) & \text{var}(X_2) & \ldots & \text{cov}(X_2, X_N) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(X_N, X_1) & \text{cov}(X_N, X_2) & \ldots & \text{var}(X_N)
\end{pmatrix}
$$

Independence and Multivariate Normal

Proposition

Suppose X and Y are independent. Then

$$cov(X, Y) = 0$$

Proof.

Suppose X and Y are independent.

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y]$$

Calculating $E[XY]$

$$\begin{aligned}
E[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} xf_X(x)dx \int_{-\infty}^{\infty} yf_Y(y)dy \\
&= E[X]E[Y]
\end{aligned}$$

Then $\text{cov}(X, Y) = 0$.

□

Zero covariance does not generally imply Independent

Suppose $X \in \{-1, 1\}$ with $P(X = 1) = P(X = -1) = 1/2$.
Suppose $Y \in \{-1, 0, 1\}$ with $Y = 0$ if $X = -1$ and $P(Y = 1) = P(Y = -1)$ if $X = 1$.

$$\begin{aligned}
E[XY] &= \sum_{i \in \{-1,1\}} \sum_{j \in \{-1,0,1\}} ijP(X = i, Y = j) \\
&= -1 \times 0 \times P(X = -1, Y = 0) + 1 \times 1 \times P(X = 1, Y = 1) \\
&\quad -1 \times 1 \times P(X = 1, Y = -1) \\
&= 0 + P(X = 1, Y = 1) - P(X = 1, Y = -1) \\
&= 0.25 - 0.25 = 0 \\
E[X] &= 0 \\
E[Y] &= 0
\end{aligned}$$

Proposition

Suppose $\boldsymbol{X} \sim$ *Multivariate Normal*$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *where* $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$. *If* $cov(X_i, X_j) = 0$, *then* X_i *and* X_j *are independent*

Tomorrow

- Changing Coordinates
- Moment Generating Functions
- Famous Inequalities
- Different Notions of Convergence