
Amy Braverman

(Jet Propulsion Laboratory)

# Basic Probability
# Part 2

Sponsored by the Jet Propulsion Laboratory and by the Keck Institute for Space Studies, California Institute of Technology

# Outline

Introduce mathematical formalism for coding and describing the outcomes of uncertain phenomena:

- Random variables.
- Distributions, densities, and mass functions.
- Expectation.

# Random variables

- A random variable (r.v.) is a numerical coding of the outcome of a trial or a set of trials.
- Example: I toss coin. $X = 1$ if it comes up heads, $X = 0$ if it comes up tails.
- Random variables can be discrete (taking on at most a countable number of values) or continuous.
- Example of a discrete r.v.: the number of times I say “hello" today.
- Example of a continuous r.v.: the height of the next person I meet.

# Random variables

Notation is very important:

- Random variables (scalars) are denoted by capital letters, e.g., $X$.
- Ordinary variables that take on fixed but possibly arbitrary values are denoted by lower-case letters, e.g., $x$. In the context of the event $\{X = x\}$, we say that $x$ is a realization of $X$.
- We may also have random vectors (a collection of random variables representing a point in high-dimensional space), and these are denoted by bold: $\mathbf{X}$
- Example of a discrete r.v.: the number of times I say “hello" today.
- Example of a continuous r.v.: the height of the next person I meet.

# Distributions, densities, and mass functions

- The behavior of a random variable is described by its cumulative distribution function (CDF):

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

- The function $P(X = x)$ is called the probability mass function (PMF) if $X$ is discrete. In this case,

$$F_X(a) = P(X \leq a) = \sum_{x \leq a} P(X = x).$$

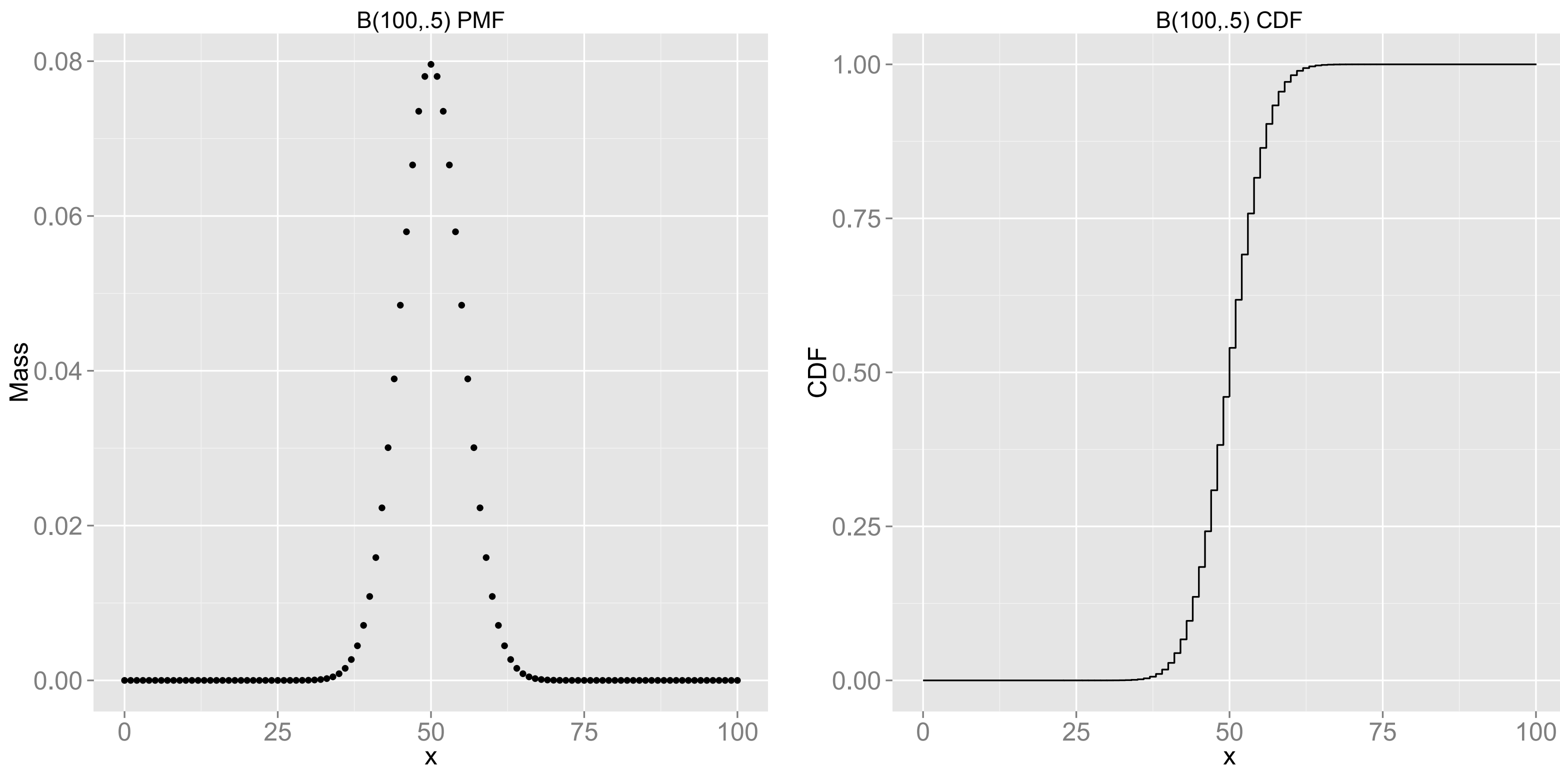
- The function $P(X = x)$ is called the probability density function (PDF) if $X$ is continuous. In this case,

$$F_X(a) = P(X \leq a) = \int_{x \leq a} f_X(x)\, dx,$$

where $f_X(x) = P(X = x)$ and $f_X(x)$ is the derivative of $F_X(x)$.

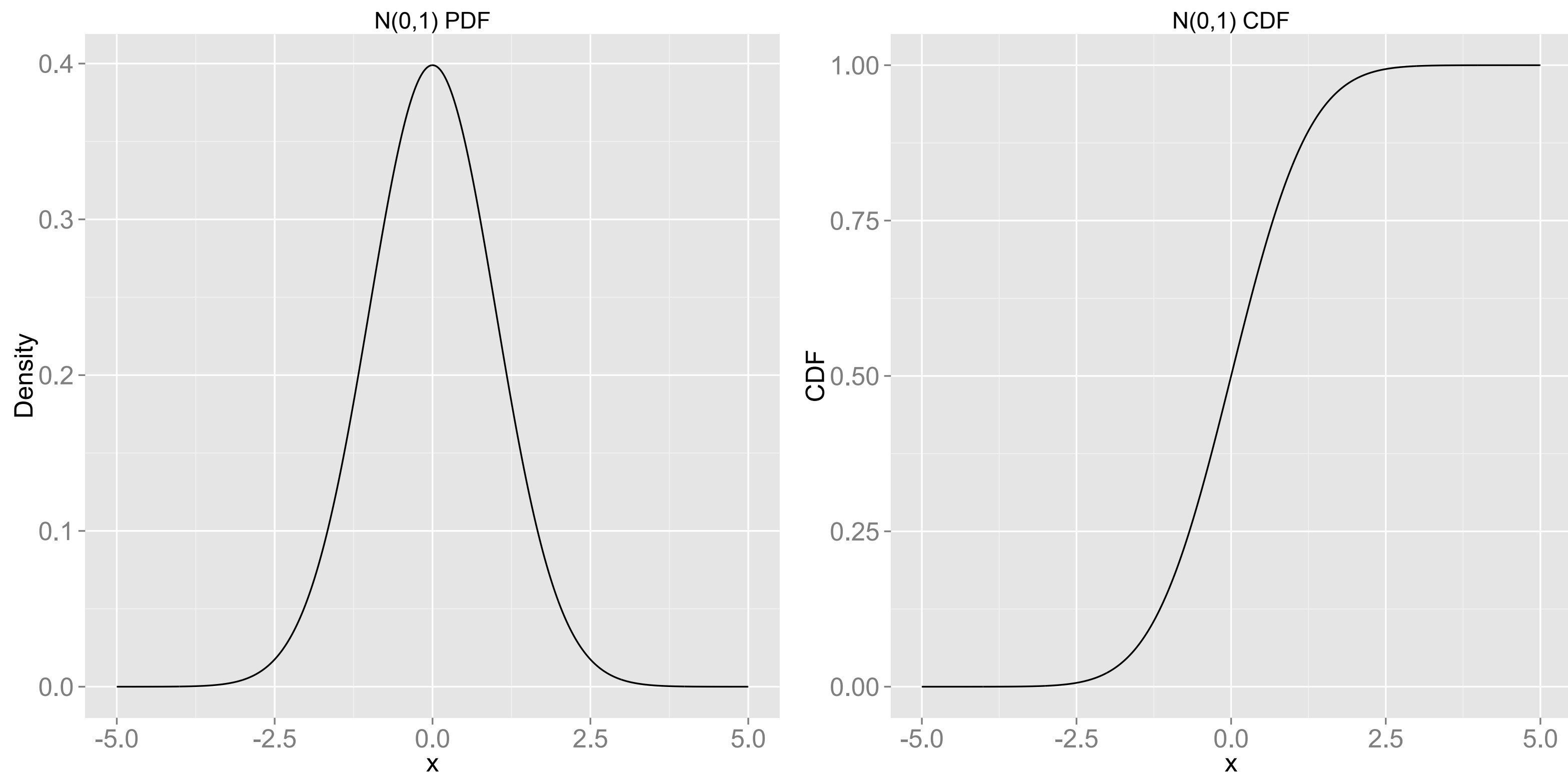
# Distributions, densities, and mass functions

## PMF and CDF of a discrete random variable:



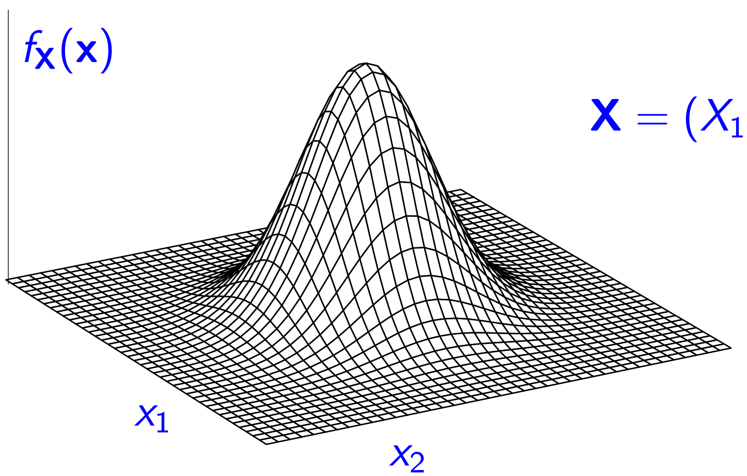
# Distributions, densities, and mass functions

## PDF and CDF of a continuous random variable:

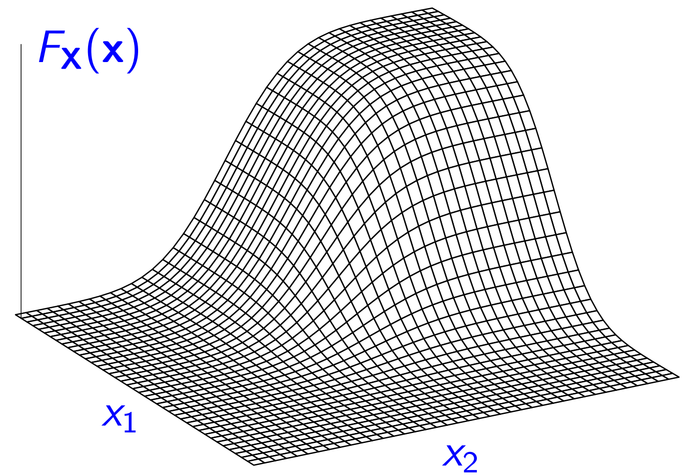


# Distributions, densities, and mass functions

## PDF and CDF of a bivariate random vector:




$$\mathbf{X} = (X_1, X_2)$$



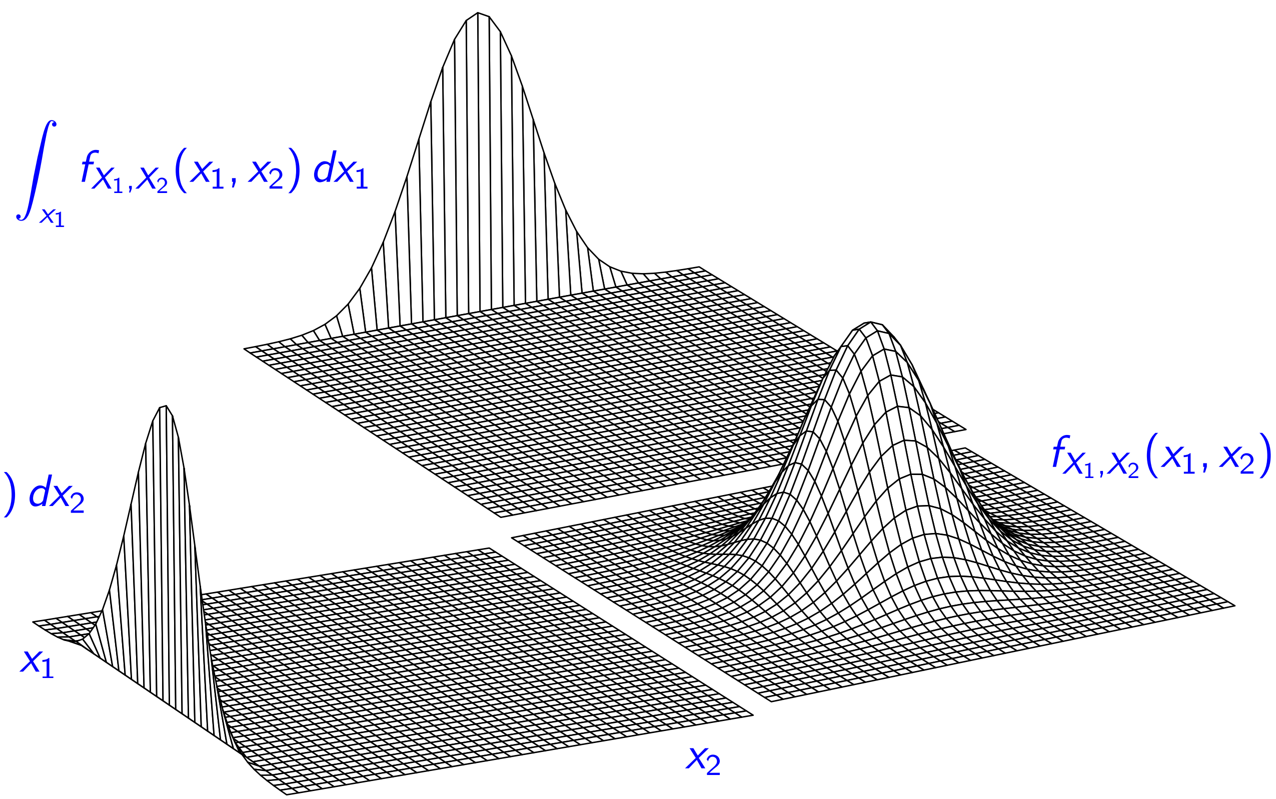

- Definitions generalize straightforwardly to higher dimensions:

$$f_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2), \quad F_{X_1,X_2}(a, b) = P(X_1 \leq a, X_2 \leq b).$$

# Distributions, densities, and mass functions

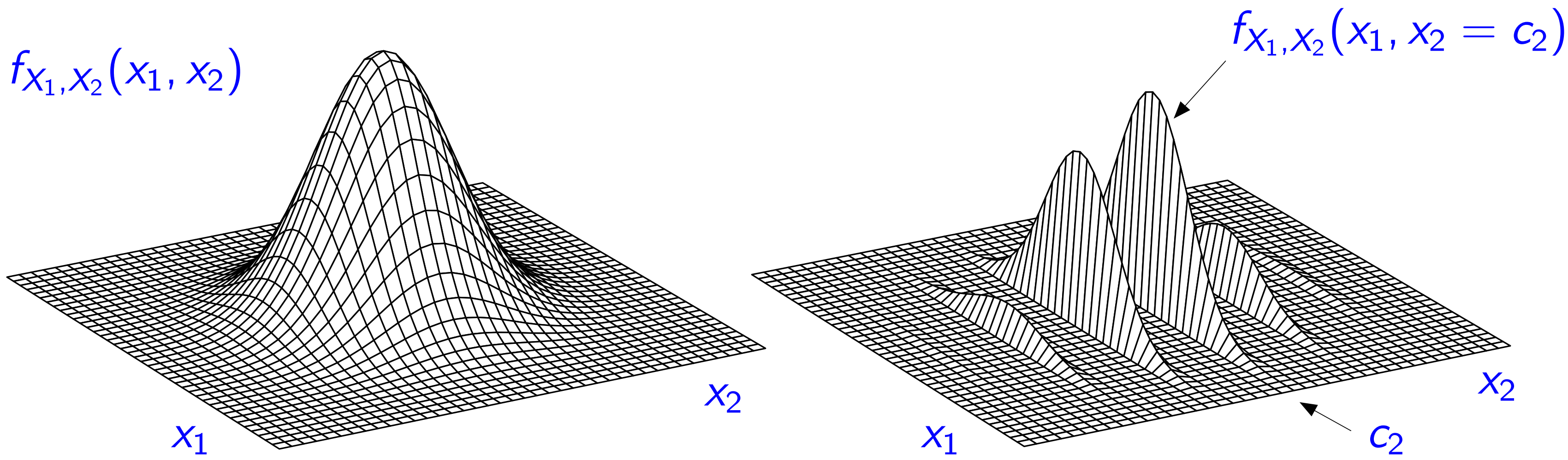
$$f_{X_2}(x_2) = \int_{x_1} f_{X_1,X_2}(x_1, x_2)\, dx_1$$

$$f_{X_1}(x_1) = \int_{x_2} f_{X_1,X_2}(x_1, x_2)\, dx_2$$




- Marginal densities: integrate a continuous joint density (or sum a discrete mass function) over the other variable (by the Law of total probability).

# Distributions, densities, and mass functions




- Conditional density: a "slice" of the joint density, renormalized to integrate to one.

$$f_{X_1|X_2}(x_1|x_2 = c_2) = \frac{f_{X_1,X_2}(x_1, x_2 = c_2)}{\int_{x_1} f_{X_1,X_2}(x_1, x_2 = c_2)\, dx_1}.$$

# Distributions, densities, and mass functions

PMF and PDF of a function of a random variable:

- Suppose $X$ is a r.v. with distribution function $F_X(x)$. What is the distribution function of $Y = g(X)$?

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

- If $X$ and $Y$ are discrete, then

$$f_Y(y) = f_X(g^{-1}(y)).$$

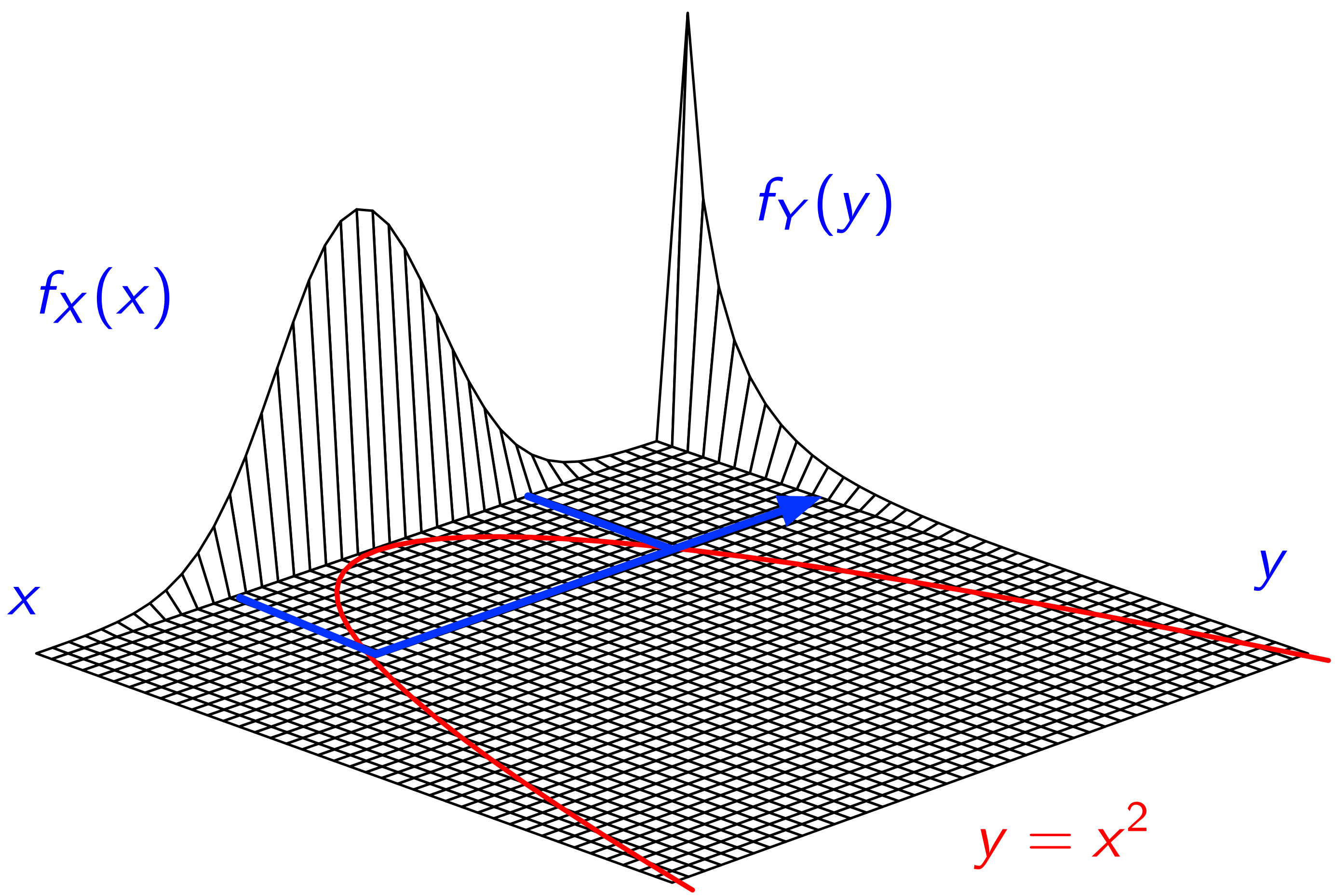
- If $X$ and $Y$ are continuous, then

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

- The joint distribution of $X$ and $Y$ is

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x).$$

# Distributions, densities, and mass functions



PDF of a function of a random vector:

- Suppose $X_1$ and $X_2$ are jointly continuous, $Y_1 = g_1(X_1, X_2)$, and $Y_2 = g_2(X_1, X_2)$.
- Suppose that $g_1(\cdot,\cdot)$ and $g_2(\cdot,\cdot)$ have continuous partial derivatives at all $(x_1, x_2)$; that there exist inverses $h_1(\cdot,\cdot)$ and $h_2(\cdot,\cdot)$ such that

$$x_1 = h_1(y_1, y_2), \quad x_2 = h_2(y_1, y_2),$$

  and that the determinant,

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} \neq 0.$$

- Then,

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) \left|J(x_1, x_2)\right|^{-1}.$$

# Expectation

- The expected value of random variable $X$ (sometimes also called the mean) is the weighted average of its potential realizations, where the weights are the probabilities associated with the realizations:

$$E(X) = \sum_x xP(X = x) \text{ (discrete)}, \qquad E(X) = \int_x xP(X = x)\, dx \text{ (continuous)},$$

or equivalently,

$$E(X) = \sum_x xf_X(x), \qquad E(X) = \int_x xf_X(x)\, dx.$$

- The expected value of a random vector is the vector of expected values of its components:

$$E(\mathbf{X}) = E(X_1, X_2)^T = (E(X_1), E(X_2))^T.$$

# Expectation

- Expected value of a function of a random variable:

$$E(g(X)) = \sum_x g(x) f_X(x) \ \text{(discrete)}, \quad E(g(X)) = \int_x g(x) f_X(x)\, dx \ \text{(continuous)}.$$

- The expected deviation of $X$ from its own expected value (mean):

$$E(X - \mu_X) = \begin{cases} \sum_x (x - \mu_X) f_X(x) & \text{discrete,} \\ \int_x (x - \mu_X) f_X(x)\, dx & \text{continuous,} \end{cases}$$

where $\mu_X = E(X)$.

- The variance of a random variable is its expected squared deviation from its mean:

$$E(X - \mu_X)^2 = \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) & \text{discrete,} \\ \int_x (x - \mu_X)^2 f_X(x)\, dx & \text{continuous.} \end{cases}$$

- The covariance of two random variables is,

$$cov(X_1, X_2) = E(X_1 - \mu_{X_1})(X_2 - \mu_{X_2}).$$

- The variance of a random vector is a matrix; the variance-covariance matrix:

$$var(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T,$$

$$= \begin{pmatrix} E(X_1 - \mu_{X_1})(X_1 - \mu_{X_1}) & E(X_1 - \mu_{X_1})(X_2 - \mu_{X_2}) \\ E(X_2 - \mu_{X_2})(X_1 - \mu_{X_1}) & E(X_2 - \mu_{X_2})(X_2 - \mu_{X_2}) \end{pmatrix}.$$

- We often use $\sigma^2_{X_i}$ as shorthand for $E(X_i - \mu_{X_i})(X_i - \mu_{X_i})$, and $\sigma_{X_i X_j}$ as shorthand for $E(X_i - \mu_{X_i})(X_j - \mu_{X_j})$.

# Expectation

- The cross-covariance between two random vectors is a matrix:

$$cov(\mathbf{X}, \mathbf{Y}) = E(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})^T,$$

$$= \begin{pmatrix} E(X_1 - \mu_{X_1})(Y_1 - \mu_{Y_1}) & E(X_1 - \mu_{X_1})(Y_2 - \mu_{Y_2}) \\ E(X_2 - \mu_{X_2})(Y_1 - \mu_{Y_1}) & E(X_2 - \mu_{X_2})(Y_2 - \mu_{Y_2}) \end{pmatrix},$$

  where **X** and **Y** are both two-dimensional here.

- In general, the cross-covariance matrix need not be square or symmetric.

# Expectation

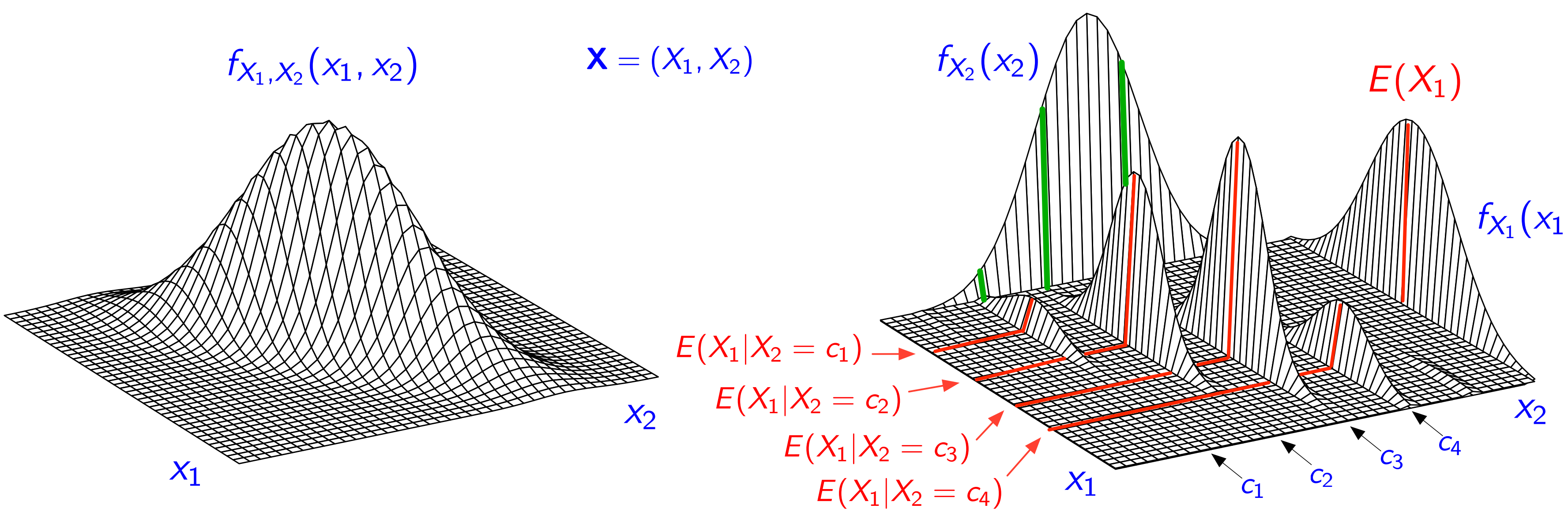
Properties of expectation for continuous r.v.'s (discrete and vector analogs are similar):

- Expectation is a linear operator: $E(a_1 X_1 + a_2 X_2) = a_1 E(X_1) + a_2 E(X_2)$.
  - Follows from $E[g(X_1, X_2)] = \int_{x_1} \int_{x_2} g(x_1, x_2) f_{X_1, X_2}(x_1, x_2)\, dx_1\, dx_2$.
  - General: $E(\sum_{i=1}^{N} a_i X_i) = \sum_{i=1}^{N} a_i E(X_i)$.
- If $X_1$ and $X_2$ are independent, then $E[g(X_1)h(X_2)] = E[g(X_1)]E[h(X_2)]$.
- $E(X_1 | X_2 = x_2) = \int_{x_1} x_1 f_{X_1 | X_2}(x_1 | x_2)\, dx_1$, viewed as a function of $x_2$, is the regression of $X_1$ on $X_2$.

# Expectation

- Law of iterated conditional expectation:

$$E[E(X_1|X_2)] = \int_{x_2} E(X_1|X_2 = x_2)\, f_{X_2}(x_2)\, dx_2 = E(X_1).$$



Properties of variance for continuous r.v.'s (discrete and vector analogs are similar):

- Variance of a linear function:

$$var(a_1X_1 + a_2X_2) = var(a_1X_1) + var(a_2X_2) + 2\,cov(a_1X_1, a_2X_2),$$
$$= a_1^2 var(X_1) + a_2^2 var(X_2) + a_1 a_2 Cov(X_1, X_2).$$
$$var\left(\sum_{i=1}^{N} a_i X_i\right) = \sum_{i=1}^{N} a_i^2 var(X_i) + 2\sum_{i=1}^{N}\sum_{j=1}^{i-1} a_i a_j cov(X_i, X_j)$$

- Variance of a non-linear function: suppose $Y = g(X) \approx g(\mu_X) + g'(\mu_X)(X - g(\mu_X))$. Then,

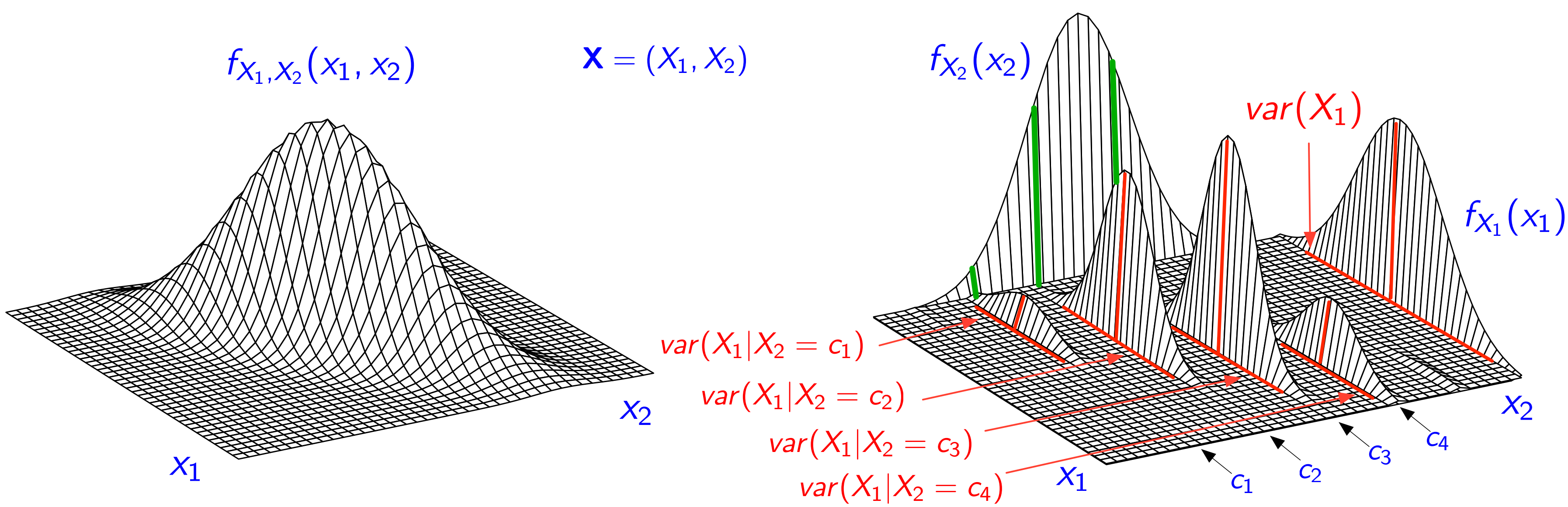
$$var(g(X)) \approx \left[g'(\mu_X)\right]^2 var(X).$$

- If $X_1$ and $X_2$ are independent, then $cov(X_1, X_2) = 0$, but the converse is not necessarily true.

# Expectation

- Conditional variance formula:

$$var(X_1) = var\left[E(X_1|X_2)\right] + E\left[var(X_1|X_2)\right].$$



# Next

By now you must be thinking, “Why should I care about this stuff?"

- We will model unknown or uncertain population characteristics with probability distributions. Call these distributions process distributions.
- To make inferences about process distributions, we compute statistics from samples.
- Statistics are themselves random variables, and have distributions of their own. Call these sampling distributions.
- The discipline of Statistics is largely concerned with understanding the relationship between a process distribution parameter of interest and the sampling distribution of a statistic designed to estimate it.

# Bibliography


- *A First Course in Probability* by Sheldon Ross, Prentice Hall, 2010.
- *An Introduction to Probability Theory and Its Applications* Volumes 1 and 2, by William Feller, John Wiley and Sons, 1957.

# Next

In the next module, we will look at how statistical inferences are made using probability models.