January 31, 2020: Probability Review

Today in recitation, we will cover:

1. Fundamentals of Probability
    - Events
    - Random Variables
    - Probability Measures

2. Probability Distributions
    - CDF, PMF, PDF
    - Mean, Mode, Median, Variance

3. Multivariate Distributions
    - Joint & Marginal Distributions
    - Conditional Probability
    - Bayes Theorem
    - Independence

Fundamentals of Probability

"Probability theory is nothing but common sense reduced to calculation." -- Pierre Laplace, 1812

Notation

X	a random variable, $X \subset \Omega$
$x \sim X$	a sample value of a random variable X
$P(x) = P(X = x)$	the probability of encountering x
$P(\bar{x}) = 1 - P(x)$	the probability of **not** encountering x
$F_X(x)$	The Cumulative Distribution Function (CDF)
$p_X(x)$	The Probability Mass Function (PMF)
$f_X(x)$	The Probability Density Function (PDF)

Definitions

Stanford CS229: Machine Learning Review of Probability Theory was used as a reference for the definitions.

Sample Space: "The set of all the outcomes of a random experiment."[1]

Events: Possible outcomes, e.g., from an experiment.

Probability Measure:

- The probability of any event A is greater than or equal to zero: $P(A) \geq 0$

- The sum of probabilities of all possible events is equal to **1**

Random Variable: a function that maps an uncertain outcome to a real number. May be discrete or continuous. Random variables are typically denoted by uppercase letters, X. Their values with lower case letters, $x \sim X$.

Probability Distributions

Cumulative Distribution Function: $F_X(x) \triangleq p(X \leq x)$

the probability the random variable takes on a value less than or equal to x. CDFs are strictly increasing.

- $0 \leq F(x) \leq 1$
- $lim_{x \to -\infty} F_X(x) = 0$
- $lim_{x \to \infty} F_X(x) = 1$
- $x \leq y \Rightarrow F_X(x) \leq F_X(y)$

Probability Mass Function: $p_X(x) \triangleq P(X = x)$

The probability a **discrete** random variable X equals x.

- $0 \leq p_X(x) \leq 1$
- $\sum_{x \in X} p_X(x) = 1$

Probability Density Function: $f_X(x) \triangleq \frac{dF_X(x)}{dx}$

For continuous random variables, the PDF is the derivative of the CDF, *if it exists*.

$f_X(x)$ is **not** the probability that $X = x$.

But the PDF can be used to compute the probability of event A: $\int_{x \in A} f_X(x)dx = P(X \in A)$.

- $f_X(x) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $f_X(x)$ may be greater than 1 for some x

Expected Value: $E[g(X)]$

If $g(X)$ is a function of random variable X, the expected value of $g(X)$ is:

for a discrete random variable:

$$E[g(x)] \triangleq \sum_{x \in X} g(x) p_X(x)$$

for a continuous random variable:

$$E[g(X)] \triangleq \int_X g(x) f_X(x) dx$$

Mean: E[X]

The expected value of the random variable.

for a discrete random variable:

$$E[X] \triangleq \sum_{x \in X} x * p_X(x)$$

for a continuous random variable:

$$E[X] \triangleq \int_X x * f_X(x) dx$$

- $E[a] = a$ for constant a
- $E[a * g(X)] = a * E[g(X)]$ for constant a
- $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$

Variance: "a measure of how concentrated the distribution of a random variable is around its mean" [1].

$$Var[X] \triangleq E[(X - E[X])^2] = E[X^2] - E[X]^2$$

- $Var[a] = 0$ for constant a
- $Var[a * g(X)] = a^2 * Var[g(X)]$ for constant a

Mode:

for a discrete random variable:

$$\triangleq argmax_x p_X(x)$$

for a continuous random variable, the modes are all local maxima of the PDF, $f_X(x)$.[2]

Median:

Let $F_X^{-1}(x)$ denote the inverse of the CDF $F_X(x)$. $F_X^{-1}(\alpha) = x_\alpha$ such that $P(X \leq x_\alpha) = \alpha$.
$F_X^{-1}(0.5)$ is the **median** of the distribution.

Multivariate Distributions

Joint Distributions reflect the probability of more than one random variable. **Marginal Distributions** consider a subset of random variables from a joint distribution. *You can marginalize out some random variables from a joint distribution.*

For two random variables:

	Joint Distribution	Marginal Distribution
CDF	$F_{XY}(x,y) = P(X \leq x, Y \leq y)$	$F_X(x) = lim_{y \to \infty} F_{XY}(x,y)$
PMF	$p_{XY}(x,y) = P(X = x, Y = y)$	$p_X(x) = \sum_{y \in Y} p_{XY}(x,y)$
PDF	$f_{XY}(x,y) = \dfrac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$	$f_X(x) = \int_Y f_{XY}(x,y) dy$

For n random variables:

	Joint Distribution	Marginal Distribution
CDF	$F_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$	$F_{X_1}(x_1) = lim_{x_2 \to \infty} \ldots lim_{x_n \to \infty} F_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_n)$
PMF	$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$	$p_{X_1}(x_1) = \sum_{x_2 \in X_2} \ldots \sum_{x_n \in X_n} p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$
PDF	$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = \dfrac{\partial^n F_{X_1,X_2,\ldots,X_N}(x_1,x_2,\ldots,x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}$	$f_{X_1}(x_1) = \int_{X_2} \ldots \int_{X_n} p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) dx_2 \ldots dx_n$

Conditional Probabilities

What is the probability a subset of random varaibles take on certain values when we know the value of

another subset of random variables?
For two random variables:

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

An analogous definitions exist for continuous random variables using PDFs.

Bayes Theorem

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

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

$$p_{X|Y}(x|y) * p_Y(y) = p_{XY}(x, y)$$

$$p_{Y|X}(y|x) * p_X(x) = p_{X|Y}(x|y) * p_Y(y)$$

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y) * p_Y(y)}{p_X(x)}$$

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

The Chain Rule
We can use conditional probabilities to rewrite a joint distribution:

$$p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = p_{X_1, X_2, \ldots, X_n}(x_n | x_1, x_2, \ldots x_{n-1}) * p_{X_1, X_2, \ldots, X_{n-1}}(x_1, x_2, \ldots x_{n-1})$$

$$p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1) \prod_{i=2}^{n} p_{X_1, \ldots, X_{i-1}}(x_i | x_1, \ldots x_{i-1})$$

Independence:

A set of events $A_1, \ldots A_n$ are **mutually independent if and only if** for any subset $S \subseteq \{1, 2, \ldots, n\}$ the probability of the events happening together is the product of the probability of each event occuring.

$$P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$$

A set of random variables are independent **if and only if**:

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}p_{X_i}(x_i)$$

- If X is independent of Y, then $p_{XY}(x|y)=p_X(x)$

Expected Values of Multivariate Distributions: $E[g(X_1,X_2,\ldots,X_n)]$

If $g(X_1,X_2,\ldots,X_n)$ is a function of random variables, the expected value of $g(X_1,X_2,\ldots,X_n)$ is:

for a discrete random variable:

$$E[g(X_1,X_2,\ldots,X_n)]\triangleq\sum_{x_1\in X_1}\cdots\sum_{x_n\in X_n}g(X_1,X_2,\ldots,X_n)*p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$$

for a continuous random variable:

$$E[g(X_1,X_2,\ldots,X_n)]\triangleq\int_{X_1}\cdots\int_{X_n}g(X_1,X_2,\ldots,X_n)*f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)\;dx_1\ldots dx_n$$

Covariance measures how (whether) two random variables vary together.

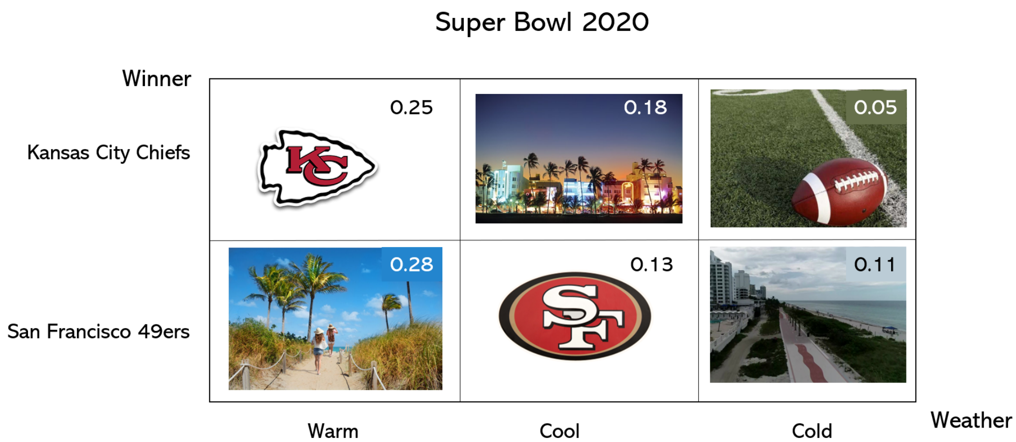
$$Cov[X,Y]\triangleq E[(X-E[X])*(Y-E[Y])]$$

$$=E[XY]-E[X]E[Y]$$

If $Cov(X,Y)=0$ then X and Y are **uncorrelated**.

Super Bowl Stats

Sample Spaces



The number in each square represents the joint probability of the specified winner and weather.

1. Do the numbers in the table represent a valid Probability Mass Function?

   Yes, all values are positive and the values sum to 1.

2. How many outcomes are in the sample space?

   6

3. How many events are in the event space?

   This is a trick question: we haven't defined any event space so we don't know. In general the number of events is not necessarily equal to the number of samples in the sample space.

Events

1. Are A and B independent?

   No. $P(A \cap B) \neq P(A) * P(B)$
   $P(A \cap B) = 0.18$
   $P(A) * P(B) = 0.1488$

2. Is $A \subset B$?

   No, some samples in A are outside B.

3. Does $A \cap B$?

   Yes.

Random Variables

- What is $p_{XY}(1, 2)$?

  0.18

- What is $F_Y(2)$?

  0.84

- What is $F_{XY}(1, 2)$?

  0.43

  The 2D CDF is shown below:

- What is $f_{XY}(1, 2)$?

N/A, our CDF is not differentiable. See the Ticket Out answer and [3] for more details.

- What is the expected value of Y?

1.63

- Does the weather impact who wins? *(Are X and Y independent?)*

Yes, the weather does impact who wins. X and Y are not independent. You can prove this by showing that $P_{XY}(x, y) \neq P_X(x) * P_Y(y)$ for any example or by showing that $P_{XY}(x|y) \neq P_{XY}(x, y) * P_Y(y)$.

- What is the covariance of X and Y?

$$Cov[X, Y] = E[XY] - E[X] * E[Y]$$

$$E[X] = 1.52$$

$$E[Y] = 1.63$$

$$E[XY] = 2.5$$

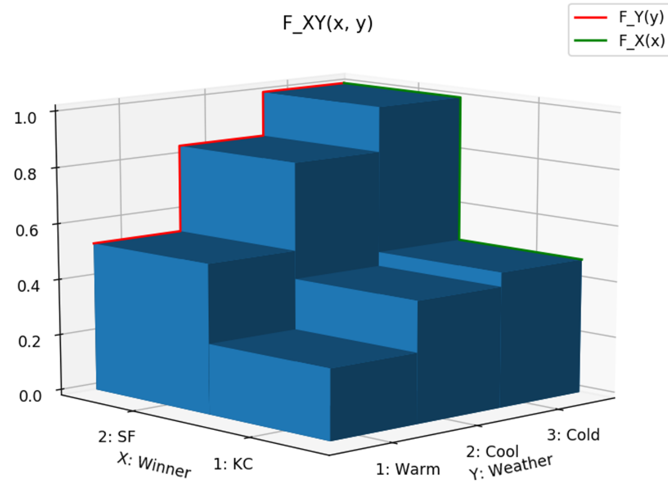
$$Cov[X, Y] = 2.5 - 1.52 * 1.63$$

$$= 0.0224$$

- If you know the 49ers will win, what is the probability that the weather will be cold?

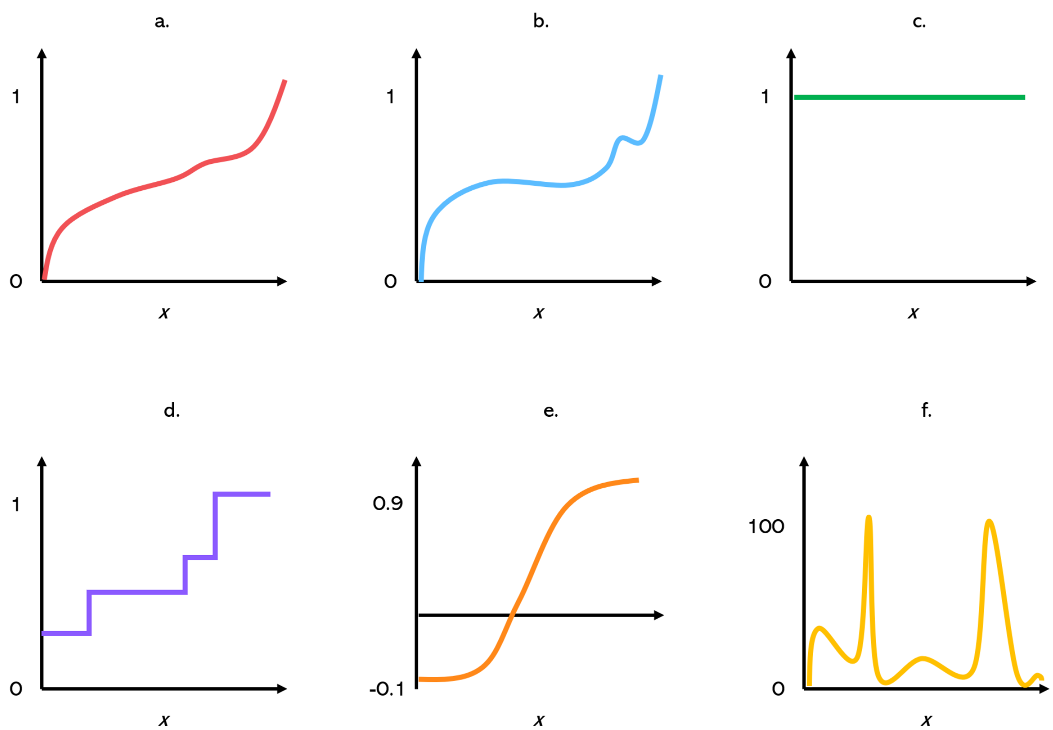
0.212

The 2D CDF of our Super Bowl example:



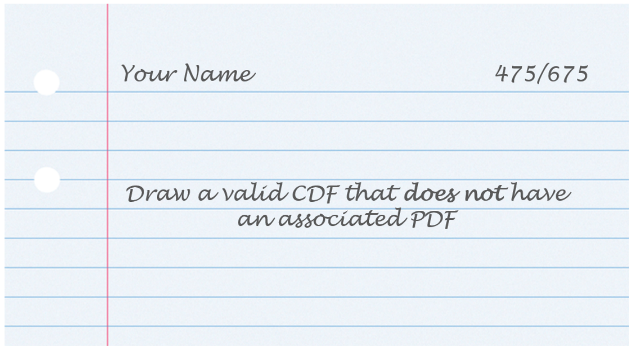
Probability Distribution Practice

Which of the following are valid Cumulative Distribution Functions?

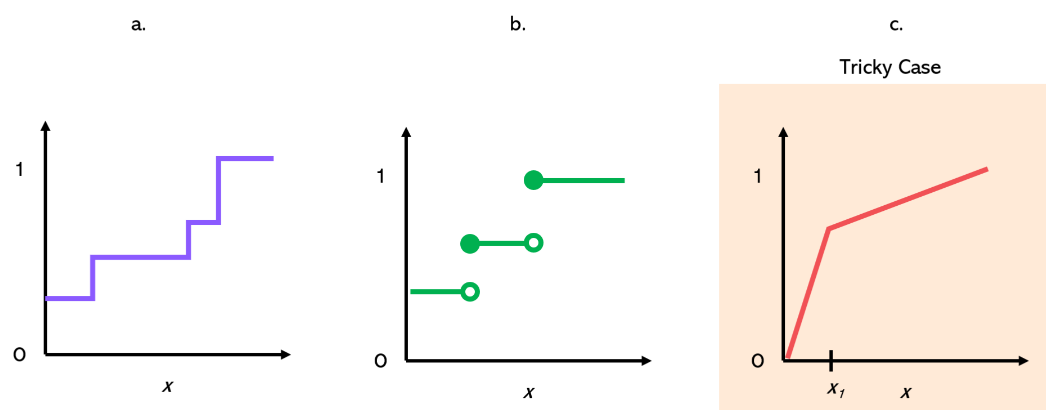


(a), (c), and (d) are valid CDFs.

Ticket Out:



(a) and (b) are correct answers. (c) is a reasonable answer, but technically we can still define a PDF for (c).



For a CDF to have a PDF, the function must be differentiable everywhere. In some cases, a function may be differentiable everywhere except a finite number of discontinuities. In these cases, the PDF may be defined everywhere except for the discontinuities, where the PDF is defined to be 0, see [3] for more details. In this course, we will still claim that **step function CDFs have no valid PDF**. Step function CDFs represent discrete random variables and should be represented with PMFs.

References

[1] Stanford CS229: Machine Learning Review of Probability Theory
[2] Wikipedia: Mode (statistics)
[3] UBC Blog discussing defining PDFs for step-function CDFs