Probability for Computer Science

Spring 2021

Lecture 11

Prof. Claire Monteleoni

Today

- Multiple discrete random variables
  - Conditional PMF
  - Conditional Expectation
  - Indep. of multiple r.v.s
- The Sample Mean
- If time: Continuous Random Variables

Conditional PMF

The conditional Probability Mass Function (PMF) of a random variable X, conditioned on an event A with P(A) > 0, is defined, for each x, as:

$$\boxed{p_{X|A}(x) = P(X = x \mid A) = \frac{P(\{X = x\} \cap A)}{P(A)}}$$

Conditional PMF

As x varies over all possible values of X, the events

$\{X = x\} \cap A$ are disjoint, and their union is A.

So, by Total Probability Theorem: $$P(A) = \sum_x P(\{X = x\} \cap A)$$

The definition of conditional PMF is:

$$p_{X|A}(x) = \frac{P(\{X = x\} \cap A)}{P(A)}$$

to verify the Normalization property of the conditional PMF:

$$\sum_x p_{X|A}(x) = \frac{\sum_x P(\{X = x\} \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

Conditional PMF

To compute the PMF of a random variable X, conditioned on an event A with P(A) > 0:

For each possible value x of X:

Collect all the possible outcomes in the event $\{X = x\} \cap A$

Sum their probabilities and normalize, by dividing by P(A), to obtain $p_{X|A}(x)$

Conditional PMF

The conditional PMF of a random variable X, conditioned on another random variable Y, is defined as:

$$p_{X|Y}(x \mid y) = P(X = x \mid Y = y)$$

$$p_{X|Y}(x \mid y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$\boxed{p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}}$$

Note: just apply the original definition, but now the event to condition on is {Y = y} (for y s.t. $p_Y(y) > 0$).

Conditional PMF

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

So from the joint PMF, we can compute conditional PMFs by normalizing the values in a particular row or column (divide by row/column total).

Y

X	0 (Winter)	1 (Spring)	2 (Summer)	3 (Fall)
0 (Brown)	0.1	0.1	0	0.2
1 (Blue)	0.05	0.05	0.1	0
2 (Green)	0	0.1	0.2	0.1

0.2/0.3 = 2/3

0

0.1/0.3 = 1/3

$$p_{X|Y}(x \mid 3) = \frac{p_{X,Y}(x,3)}{p_Y(3)}$$

Conditional PMF

Conditional PMFs of one random variable conditioned on another r.v. provide ways to calculate the joint PMF:

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

so $\boxed{p_{X,Y}(x,y) = p_Y(y)p_{X|Y}(x \mid y)}$ by Multiplication Rule.

And $\boxed{p_{X,Y}(x,y) = p_X(x)p_{Y|X}(y \mid x)}$ by Multiplication Rule,

and definition of $p_{Y|X}(y|x)$.

Conditional PMF

Conditional PMFs of one random variable conditioned on another r.v. provide ways to calculate the marginal PMFs:

$$\boxed{\begin{aligned} p_X(x) &= \sum_y p_Y(y) p_{X|Y}(x|y) \\[2em] p_Y(y) &= \sum_x p_X(x) p_{Y|X}(y|x) \end{aligned}}$$

by Total Probability Theorem.

Conditional Expectation

- The conditional expectation of r.v. X, given an event A with P(A)>0 is defined as: $$E[X|A] = \sum_x x \cdot p_{X|A}(x)$$
- For a function g(X), $$E[g(X)|A] = \sum_x g(x) \cdot p_{X|A}(x)$$
- Given r.v.s X and Y associated with the same experiment, the conditional expectation of X given a value y of Y is:

$$E[X|Y = y] = \sum_x x \cdot p_{X|Y}(x|y)$$

Total Expectation Theorem

- Given disjoint events $A_1, \ldots, A_n$ that partition the sample space, with $P(A_i) > 0$ for all i,

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$$

Example

Messages are sent from a computer in Boston, over the internet to the following destinations, with the following probabilities:

- NYC with probability 0.5
- DC with probability 0.3
- Denver with probability 0.2

The transit time is a random variable, T. Its expectation, conditioned on each city, is:

- 0.05 if message destination is NYC
- 0.1 if message destination is DC
- 0.3 if message destination is Denver

Q: What is E[T]?

Independence of a r.v. from an event

A random variable, X, is independent of an event, A, if, for all x,

$$P(X = x \text{ and } A) = P(X = x)P(A)$$

$$= p_X(x)P(A)$$

i.e., X is independent of A if the events {X=x} and A are independent, for every value of x.

Independence of random variables

To prove or disprove that r.v.s X and Y are independent, it is enough to prove or disprove any of the following statements (as they are equivalent):

- $p_{X,Y}(a,b) = p_X(a)p_Y(b)$ *for all a and b*
- $p_X(a) = p_{X|Y}(a \mid b)$ *for all a and b s.t.* $p_Y(b) > 0$
- $p_Y(b) = p_{Y|X}(b \mid a)$ *for all a and b s.t.* $p_X(a) > 0$

Independence of random variables

Are X and Y independent?

Y

X

	y_1	y_2	y_3	y_4
x_1	0.05	0.15	0	0.2
x_2	0.025	0.075	0	0.1
x_3	0.05	0.15	0	0.2

Independence of random variables

Are X and Y independent? Yes.

How can we tell?

Y

X		y_1	y_2	y_3	y_4
	x_1	0.05	0.15	0	0.2
	x_2	0.025	0.075	0	0.1
	x_3	0.05	0.15	0	0.2

Independence of random variables

How can we tell that X and Y are independent?

The columns are multiples of each other. Therefore $p_{X|Y}(x \mid y)$ is the same for every value of y, and therefore does not depend on y, so $p_{X|Y}(x \mid y) = p_X(x)$.

Y

X

	y_1	y_2	y_3	y_4
x_1	0.05	0.15	0	0.2
x_2	0.025	0.075	0	0.1
x_3	0.05	0.15	0	0.2

Conditional independence of r.v.s

Random variables X and Y are conditionally independent, given event A if, for all x, y,

$$P(X = x, Y = y \mid A) = P(X = x \mid A)P(Y = y \mid A)$$

$$= p_{X|A}(x)p_{Y|A}(y)$$

Equivalently, for all x and y s.t. $p_{Y|A}(y) > 0$,

$$p_{X|Y,A}(x|y) = p_{X|A}(x)$$

Properties of independent r.v.s

If X and Y are independent random variables, then:

- $E[XY] = E[X]E[Y]$
- $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

NOTE: Not true in general for arbitrary r.v.s!

Independence of multiple r.v.s

- Random variables X, Y, and Z are independent if:

  For all x, y, z:

$$p_{X,Y,Z}(x,y,z) = p_X(x)p_Y(y)p_Z(z)$$

- Let X_1 ,…, X_n be independent random variables. Then:

$$\mathtt{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathtt{var}(X_i)$$

Note: These properties need not hold in general – need independence.

The Sample Mean

- Suppose we want to estimate the approval rating of a public figure, B.
- We ask *n* people drawn uniformly at random from the population.
- Define X_i as an indicator random variable for whether the i-th person approves of B.
- We model X_1, X_2, …, X_n as independent Bernoulli random variables, with common mean, *p*, and variance *p(1-p)*.
- That is, we assume *p* is the true approval rating of B. It is unknown, so we try to estimate it.
- We compute the Sample Mean from the *n* responses, i.e. the average approval rating in the *n*-person sample:

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Mean and Variance of the Sample Mean

- The Sample Mean is:

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- What are $E[S_n]$ and $\mathrm{var}(S_n)$?

Continuous Random Variables

- Probability Density Function (PDF)
- Expectation, Variance
- Random variables: Uniform, Exponential
- Conditional Distribution Function (CDF)

Continuous random variables

A random variable, X, is continous if there is a nonnegative function f_X, called the probability density function (PDF), such that for every subset of the real line, $B \subseteq \mathbb{R}$,

$$\boxed{P(X \in B) = \int_B f_X(x)dx}$$

The probability that X falls in interval [a, b] is:

$$\boxed{P(a \le X \le b) = \int_a^b f_X(x)dx}$$

Properties of PDF

- Nonnegativity: $f_X(x) \geq 0 \qquad \forall x \in \mathbb{R}$

- Normalization:

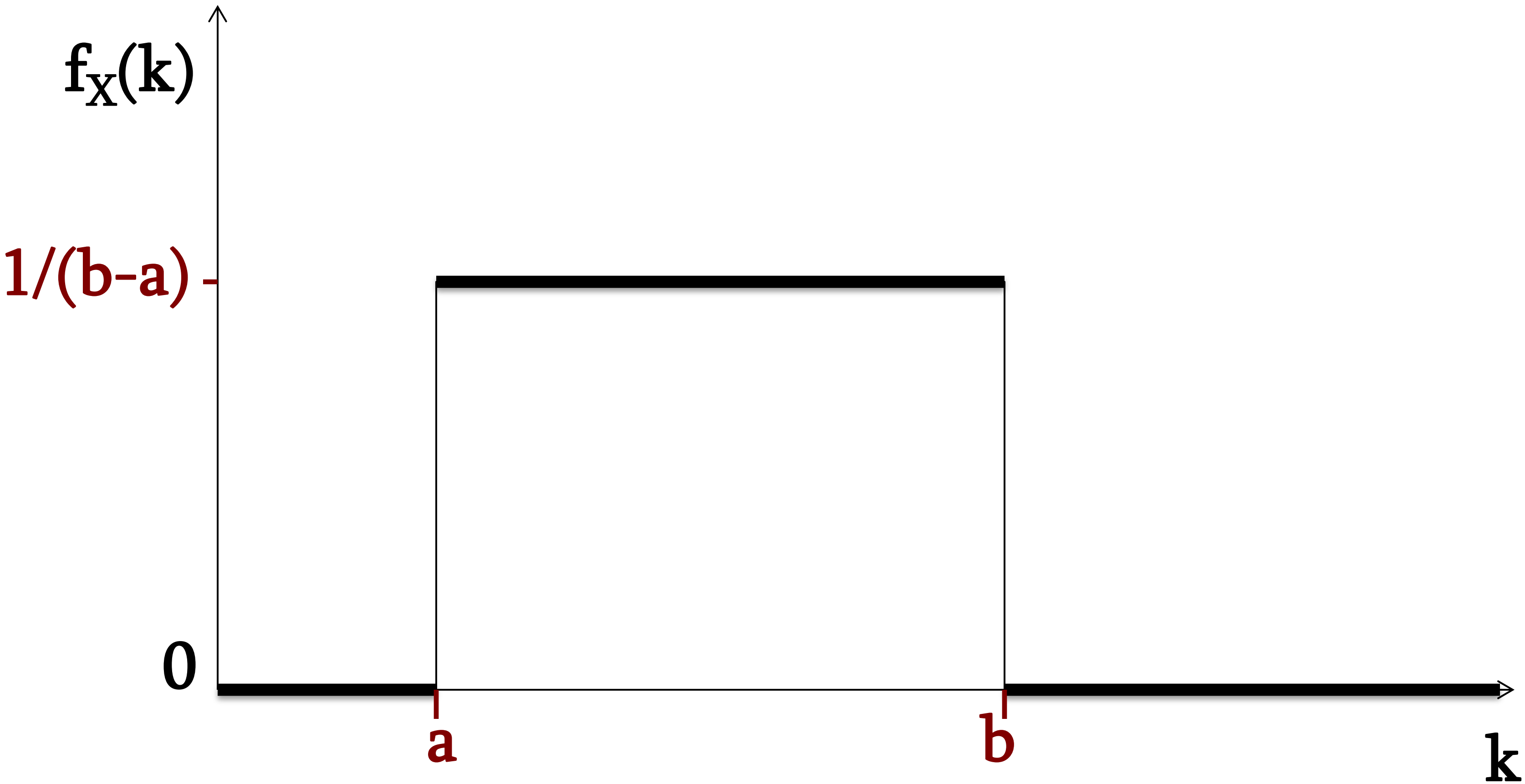
$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(x)dx = 1$$

Uniform random variable

A continuous r.v., X, is uniform (or uniformly distributed) if it takes values in an interval [a, b], and any two sub-intervals of the same length have the same probability. Its PDF is:

$$f_X(k) = \begin{cases} \frac{1}{b-a} & a \leq k \leq b \\ 0 & \text{otherwise} \end{cases}$$

Uniform PDF



Expectation of a continuous r.v.

If X is a continuous r.v., its expectation (or mean, or expected value) is defined as:

$$\boxed{E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx}$$

Note: this is exactly the center of gravity of the PDF. That is, it is the exact tipping point of the area under the PDF.

[True in the discrete case as well].

Functions of a continuous r.v.

If X is a continuous random variable, and Y is a real-valued function of X, g(X), then Y is a random variable.

- Y can be continuous, e.g. Y = 3 X.
- Y can be discrete, e.g. $Y = \begin{cases} 1 & X > 0 \\ 0 & \text{otherwise} \end{cases}$

Expected value of a function of a continuous r.v.:

$$E[Y] = \boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx}$$

Variance of continuous r.v.

Variance of a continuous r.v., X, is defined:

$$\mathrm{var}(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

Note: same definition (LHS) as in discrete case, but for the expectation of a function of a continuous r.v., need to integrate.

Similar to discrete case, n-th moment is $E[X^n]$, and:

$$0 \leq \mathrm{var}(X) = E[X^2] - (E[X])^2$$

Linear functions of a continuous r.v.

If X is a continuous r.v. and Y is a linear function of X, i.e. $Y = a X + b$ for some constants a, b, then:

1. $E[Y] = E[a X + b] = a E[X] + b$

2. $\mathrm{var}(Y) = \mathrm{var}(a X + b) = a^2 \mathrm{var}(X)$

Note: just as in discrete case.

Exponential random variable

An exponential random variable, X, is a continuous r.v. with the following PDF, for a constant λ > 0.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Exponential PDF

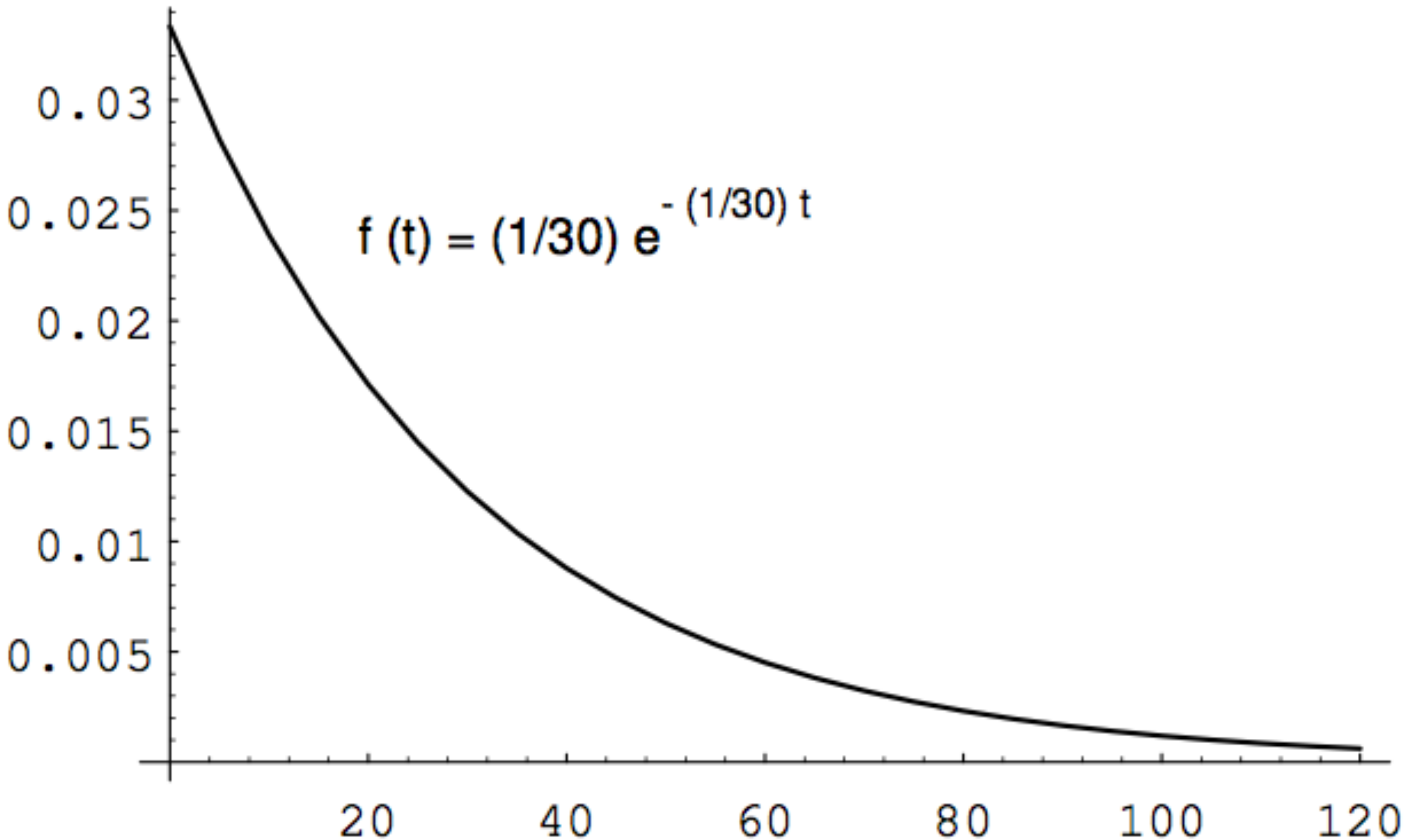


Figure 2.20: Exponential density with $\lambda = 1/30$.

Credit: Grinstead and Snell, 1997.

Cumulative distribution function (CDF)

For any r.v., X, discrete or continuous, the cumulative distribution function, CDF is:

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

which is computed as follows:

$$F_X(x) = \begin{cases} \sum_{k \leq x} p_X(k), & \text{if X is a discrete r.v.} \\ \\ \int_{-\infty}^{x} f_X(t)dt, & \text{if X is a continuous r.v.} \end{cases}$$

Compute PMF from CDF, for discrete r.v.

If X is a discrete r.v., then CDF of X is:

$$F_X(x) = \sum_{k \leq x} p_X(k)$$

Can therefore compute PMF from CDF as follows:

$$p_X(k) = P(X = k) = P(X \leq k) - P(X \leq k - 1)$$

$$= F_X(k) - F_X(k - 1)$$

Compute PDF from CDF, for continuous r.v.

If X is a continuous r.v., then CDF of X is:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

In other words, the value of the CDF at x, $F_X(x)$, is equal to the area under PDF, f_X, from $-\infty$ to x.

Can therefore compute PMF by differentiating the CDF:

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

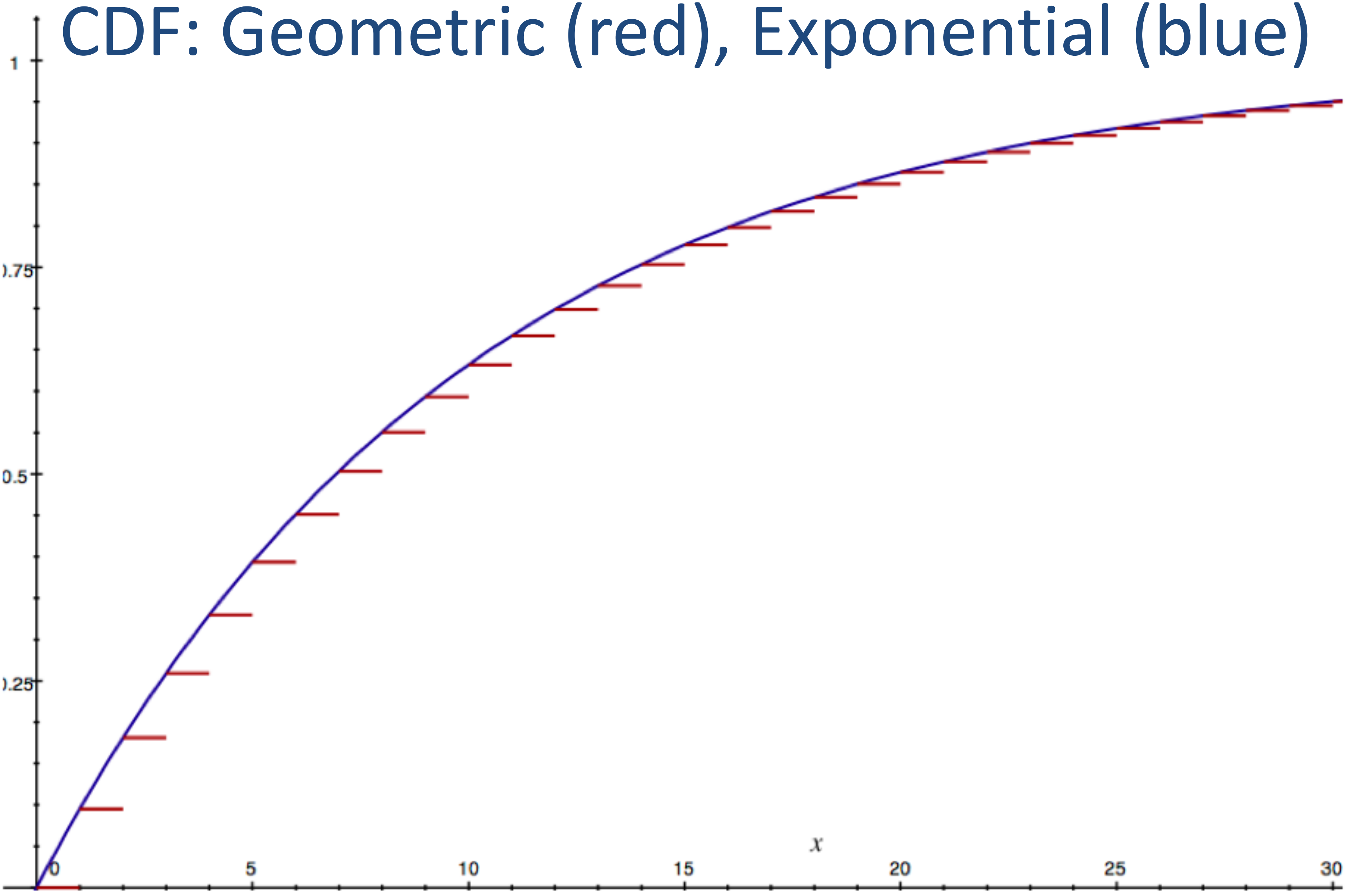
Properties of the CDF

- F_X is monotonically non-decreasing:
  if $j \leq k$ then $F_X(j) \leq F_X(k)$

- $F_X(k)$ tends to 0 as $k \rightarrow -\infty$
- $F_X(k)$ tends to 1 as $k \rightarrow \infty$

- If X is a discrete r.v. then its CDF, $F_X(k)$, is a piecewise constant function of k.
- If X is a continuous r.v. then its CDF, $F_X(k)$, is a continuous function of k.

CDF: Geometric (red), Exponential (blue)



Credit: Vaughan, UCLA, 2012.