CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 10:

- Cumulative Distribution Functions
- Standard Deviations
  - Bernoulli
  - Binomial
  - Geometric

Discrete RandomVariables: Variance

Ok, finally, here is the best definition:

$$Var(X) =_{def} E[\,(X - \mu_X)^2\,]$$

Alternate notation for expected value:

$$\mu_X = E(X)$$

or just μ if **X** is obvious.

This is the standard definition and has several advantages:

- It it much easier to work with mathematically;
- Like the absolute value, it gives only positive values.

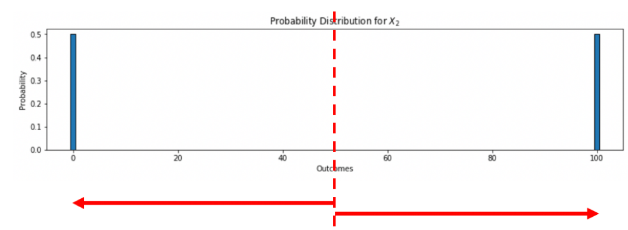
But it gives results which are not very intuitive!

$R_{X_1} = \{\, 0,\ 1 \,\}$ $\quad f_{X_1} = \{\, \frac{1}{2}, \frac{1}{2} \,\}$ $\quad E(X) = 0.5$

$R_{X_2} = \{\, 0,\ 100 \,\}$ $\quad f_{X_2} = \{\, \frac{1}{2}, \frac{1}{2} \,\}$ $\quad E(X) = 50$

$R_{(X_1-0.5)^2} = \{\, 0.25 \,\}$ $\quad f_{(X_1-0.5)^2} = \{\, 1.0 \,\}$ $\quad E[\,(X_1 - 0.5)^2] = \underline{0.25}$

$R_{(X_2-50)^2} = \{\, 2500 \,\}$ $\quad f_{(X_2-50)^2} = \{\, 1.0 \,\}$ $\quad E[\,(X_2 - 50)^2\,] = \underline{2500}$



And what about the units?
If these are dollars, then this is 2500 squared dollars...

Discrete Random Variables: Standard Deviation

Therefore a more common measure of spread around the mean is the
Standard Deviation:

$$\sigma_X =_{def} \sqrt{Var(X)}$$

$$R_{X_1} = \{ 0, 1 \} \qquad f_{X_1} = \{ \tfrac{1}{2}, \tfrac{1}{2} \} \qquad E(X) = 0.5$$

$$R_{X_2} = \{ 0, 100 \} \quad f_{X_2} = \{ \tfrac{1}{2}, \tfrac{1}{2} \} \qquad E(X) = 50$$

$$R_{(X_1-0.5)^2} = \{ 0.25 \} \qquad f_{(X_1-0.5)^2} = \{ 1.0 \} \qquad Var(X_1) = 0.25 \qquad \sigma_{X_1} = 0.5$$

$$R_{(X_2-50)^2} = \{ 2500 \} \qquad f_{(X_2-50)^2} = \{ 1.0 \} \qquad Var(X_2) = 2500 \qquad \sigma_{X_2} = 50$$

This has all the advantages of the variance, plus two more:

- The units are correct; and
- It corresponds to a well-known geometric notion, the Euclidean Distance....

Solution:

Suppose X = "the number of heads showing on a (possibly) unfair coin, where the probability of heads is p".

$$R_x = \{0, 1\}$$

$$f_x = \{(1-p), p\}$$

$$E(X) = 0 \cdot (1-p) + 1 \cdot p = p$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= (0^2 * (1-p) + 1^2 * p) - p^2 \\ &= p - p^2 \\ &= (1-p) \cdot p \end{aligned}$$

Cumulative Distribution Functions

One more topic before considering the standard distributions....

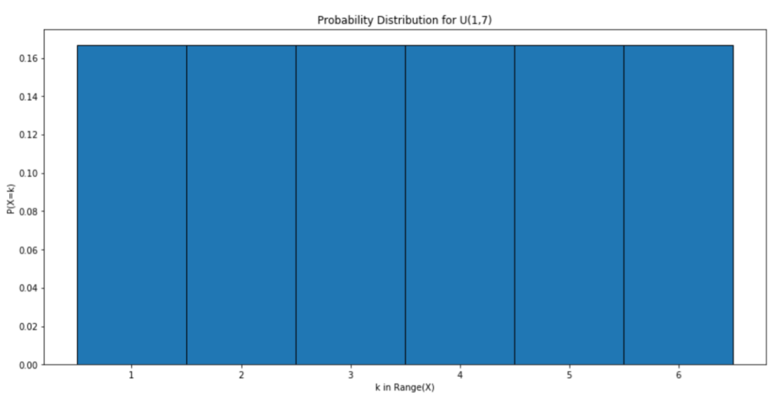
The Cumulative Distribution Function (CDF) for a random variable X shows what happens when we keep track of the sum of the probability distribution from left to right over its range:

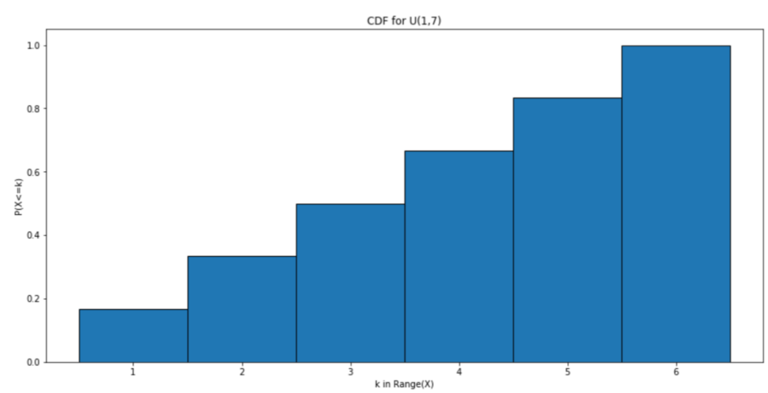
$$F_X(k) = P(X \leq k) = \sum_{a \leq k} f_X(a)$$

Example: X = "The number of dots showng on a thrown die"

Probability Distribution Function f_X

Cumulative Distribution Function F_X





Standard Distributions

Any random variable X has a distribution which can be characterized by

- R_X -- The range of the random variable
- f_X -- The Probability Distribution/Density Function PDF
- F_X -- The Cumulative Distribution Function (CDF)
- $E(X)$ -- Expected value
- $Var(X)$ -- Variance
- σ_X -- Standard Deviation

In addition, we are interested in

- Formulae for calculating f_X and F_X, if such exist (hopefully efficient!)
- Algorithms for generating random variates (as in lab!)
- Any special properties of the distribution (e.g., the "memoryless property")
- Applications (random experiments which follow that distribution)

Standard Distributions

We will look at the following distributions in the next week or so:

Discrete Distributions

- Bernoulli
- Binomial
- Geometric
- Negative Binomial
- Poisson

Continuous Distributions

- Normal
- Exponential

These are summarized, with useful code for displaying the PMF and CDF in the notebook **Distributions.ipynb** on the class web site. Wikipedia has very good pages on all these distributions. Your textbook is a little spotty on organizing this material.

Bernoulli Distribution

Suppose you have a coin where the probability of a heads is p and we define the random variable

X = "the number of heads showing on a flipped coin"

Then we say that X is distributed according to the Bernoulli Distribution with parameter p, and write this as:

$$X \sim \text{Bernoulli}(p)$$

where

$$R_X = \{0, 1\}$$
$$f_X = \{1 - p, p\}$$

and where

$$E(X) = p$$
$$Var(X) = p - p^2 = (1-p)p$$
$$\sigma_X = \sqrt{(1-p)p}$$

Jacob Bernoulli

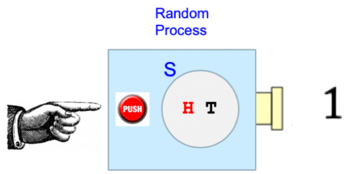
Jacob Bernoulli

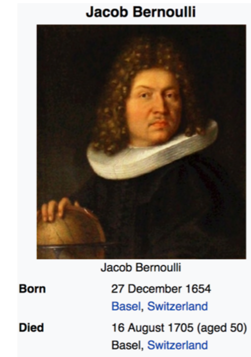
Born	27 December 1654 Basel, Switzerland
Died	16 August 1705 (aged 50) Basel, Switzerland

Among other accomplishments, Bernoulli discovered the number ***e*** (but Euler got the credit for "Euler's Number").

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$$

Bernoulli Distribution





Each "poke" of such a random variable is called a Bernoulli Trial, and the outcomes are often labelled as

1 = Success 0 = Failure

$$X \sim \text{Bernoulli}(p)$$

Bernoulli Trials, and Bernoulli random variables, describe simple random experiments where there are two possible outcomes, and the probability of the outcomes is fixed by p and (1-p); the notion of "success" and "failure" is just a convenience and does not always correspond to the desirability of the outcome:

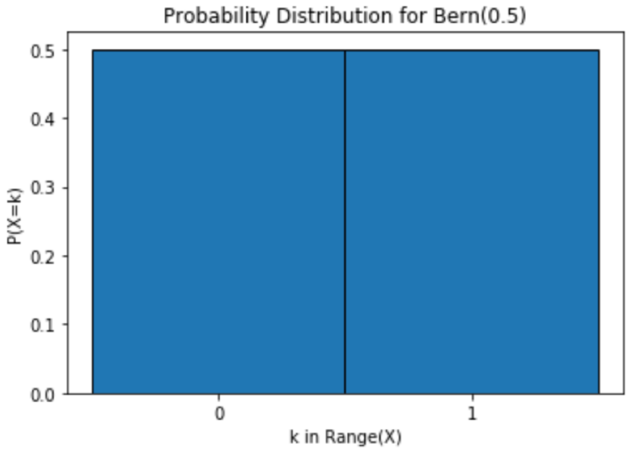
$$R_X = \{0, 1\}$$
$$f_X = \{1-p, p\}$$

$$E(X) = p$$
$$Var(X) = p - p^2 = (1-p)p$$
$$\sigma_X = \sqrt{(1-p)p}$$

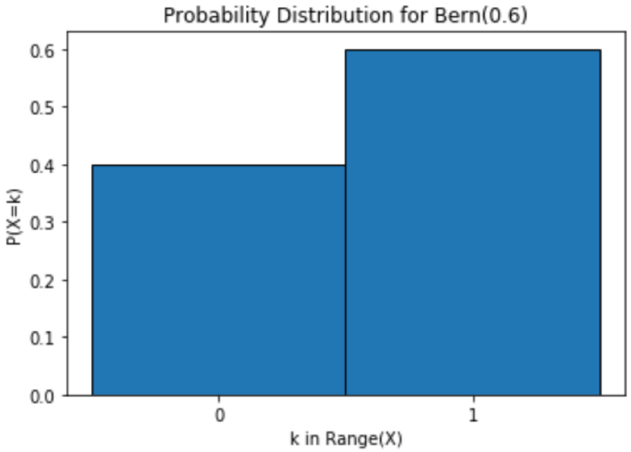
- Will I pass this course or not?
- Am I pregnant or not?
- Do I have cancer or not?

Bernoulli Distribution

The probability distribution is easy to visualize:




```
Rng(X) = [0, 1]
  f(X) = [0.5, 0.5]
```




```
Rng(X) = [0, 1]
  f(X) = [0.4, 0.6]
```

$$X \sim \text{Bernoulli}(p)$$

$$R_X = \{0, 1\}$$
$$f_X = \{1 - p, p\}$$

$$E(X) = p$$
$$Var(X) = p - p^2 = (1 - p)p$$
$$\sigma_X = \sqrt{(1 - p)p}$$

Not much more to say about this, except as a foundation for more complicated distributions such as the Binomial (also invented by Bernoulli)....

Binomial Distribution

The Binomial Distribution occurs when you count the number of successes in N independent and identically distributed Bernoulli Trials (i.e., p is the same each time).

Formally, if Y ~ Bernoulli(p), and

$$X = \text{“The number of successes in N independent trials of Y”} = \overbrace{Y + Y + \ldots + Y}^{\text{N times}}$$

then we say that X is distributed according to the Binomial Distribution with parameters N and p, and write this as:

$$X \sim B(N, p)$$

where

$$R_X = \{\, 0, \ldots, N \,\}$$

$$f_X(k) = \binom{N}{k} p^k \,(1-p)^{n-k}$$

Note: k successes and N-k failures:

SSS…FFFF…

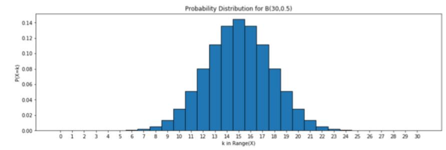
has probability $p^k \, (1\text{-}p)^{N\text{-}k}$

and there are C(N,k) such sequences.

and where

$$E(X) = N\,p \qquad Var(X) = N\,(1-p)\,p \qquad \sigma_X = \sqrt{N(1-p)p}$$

Binomial Distribution



For the visual display of this distribution, we will look briefly at the Distributions notebook to get a sense for this.....

The motivation for this distribution comes from the fact that many complex phenomena are composed of the additive effect of many small binary choices or events (Bernoulli Trials!); a vivid illustration of this can be seen in the Galton Board or Quicunx:

https://www.mathsisfun.com/data/quincunx.html

https://www.youtube.com/watch?v=J7AGOptcR1E

Phenomena explained by the binomial are widespread throughout ordinary life, biology, engineering, and business:

- You go through 10 traffic lights; what is the probability that you stop at 4 of them?
- The probability of any individual in this class having a tattoo is 0.2; what is the probability that at least 40 people have a tattoo?
- Suppose you have not studied for a multiple-choice test and you randomly guess at each problem; what is the probability that you pass the test?
- Suppose 5% of tax returns are submitted with fraudulent data and the IRS examines 1% of returns; what is the probability that they will detect 3% of all fraudulent returns?

Binomial Distribution

The binomial distribution is of widespread applicability, but it has a disadvantage: the only way to compute probabilities is to use the formula

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

and this can involve some very large numbers.... for example, $\binom{10,000}{5,000}$ is:

Binomial Distribution

The binomial distribution is of widespread applicability, but it has a disadvantage: the only way to compute probabilities is to use the formula

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

and this can involve some very large numbers.... for example, $\binom{10,000}{5,000}$ is:

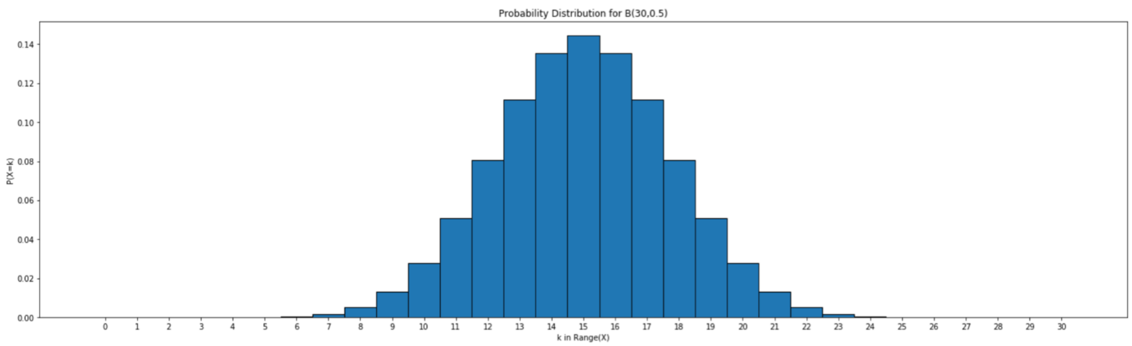
25306860788368395425009996518100609077068723635393882065500412722438406556357
59019339192864845058090054648588847516373207436678575382395143428849429230095
48452463962038917241970441315347251288293140458085203641975429015396451120702
90054172676007088840418254863094094181054099386605927195740729486409692056617
18423204598807405832967747589281702974353796541551881065990864294962746741105
93266103962513789521968236641998601608992292111723427844080079206219850932895
94893598742945434967990132009169690766577216007716147870782048492669875259433
87372044495991001950014092399556127332333759349165922429479347018286914614395
87257149876168747889795948911236702108322879910187970022052679201806406384025
98651617475613895716421284464721781536343381888676773053733597398324984189651
96131689184732588697226168384517309056386524294344846582784027540865916271808
12730710494080

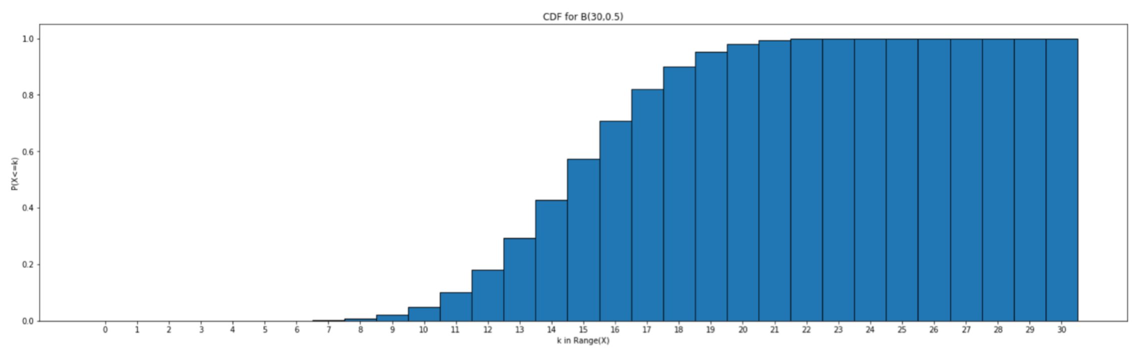
or about 10^{861}. For comparison, there are about 10^{87} atoms in the universe....

Binomial Distribution

The CDF also has no easy, efficient formula, you can only sum up the numbers involved:

$$P(X \leq k) = \sum_{a \leq k} \binom{N}{a} p^a (1-p)^{N-a}$$





Therefore, it has always been important to find more efficient ways to compute the binomial; we will see that several results in the history of probability were motivated by the desire to find easy-to-compute approximations to the binomial.

Geometric Distribution

The Geometric Distribution occurs when you count the number of independent and identically distributed Bernoulli trials until the first success.

Formally, if Y ~ Bernoulli(p), and

X = "The number of trials of Y until the first success"

then we say that X is distributed according to the Geometric Distribution with parameter p, and write this as:

$$X \sim G(p)$$

where

$$\begin{array}{rlcccccccl}
R_X = & \{ & 1, & 2, & 3, & \ldots & k, & \ldots & \} \\
S = & \{ & S, & FS, & FFS, & \ldots & FFF..S, & \ldots & \} \\
f_X = & \{ & p, & (1-p)p, & (1-p)^2 p, & \ldots & (1-p)^{k-1} p, & \ldots & \}
\end{array}$$

For k, we have k-1 failures and 1 success (FFF... S), which has probability $(1-p)^{k-1}$ p .

Geometric Distribution

Example

An absent-minded professor has 6 keys on his key ring and does not always remember which of his keys opens his office door. He chooses keys randomly and with replacement to try to open his door.

What is the probability that he opens it after only 3 tries?

Solution. This is G(1/6).

$P(X=3) = (5/6)^2\,(1/6) = 0.1157$

Geometric Distribution

$$
\begin{array}{lcccccc}
R_X = \{ & 1, & 2, & 3, & \ldots & k, & \ldots \} \\
S = \{ & S, & FS, & FFS, & \ldots & FFF..S, & \ldots \} \\
f_X = \{ & p, & (1-p)p, & (1-p)^2 p, & \ldots & (1-p)^{k-1} p, & \ldots \}
\end{array}
$$

But how do we know this is even a distribution? The only question is: Does f_X sum to 1.0?

Yes, no worries.... Suppose α is the sum of f_X. Then:

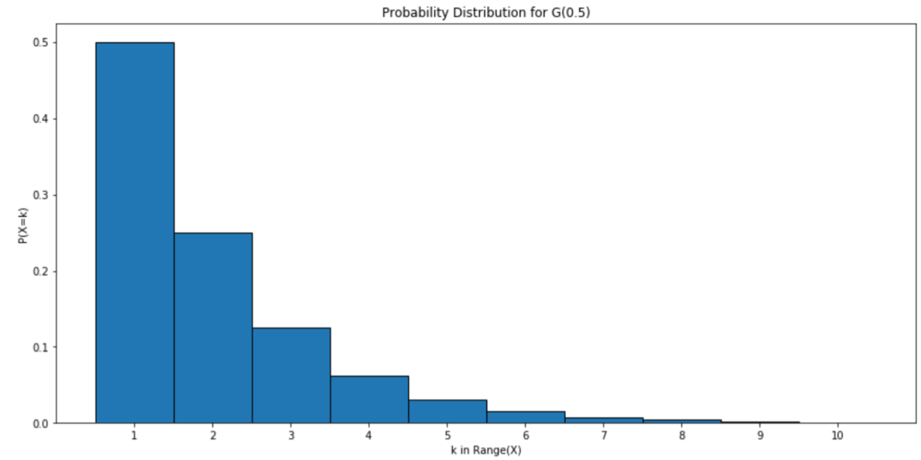
$$
\begin{aligned}
\alpha &= p + (1-p)\,p + (1-p)^2\,p + (1-p)^3\,p + \ldots \\
&= p + (1-p)\,(\,p + (1-p)\,p + (1-p)^2\,p + \ldots\,) \\
&= p + (1-p)\,\alpha \\
&= p + \alpha - p\alpha
\end{aligned}
$$

Subtracting α from both sides we have:

$$
\begin{aligned}
& 0 = p - p\alpha \\
\Leftrightarrow \quad & p\alpha = p \\
\Leftrightarrow \quad & \alpha = p/p \ = \ 1.0
\end{aligned}
$$

Geometric Distribution

Fortunately, the probability mass function and the CDF are both easy to compute:



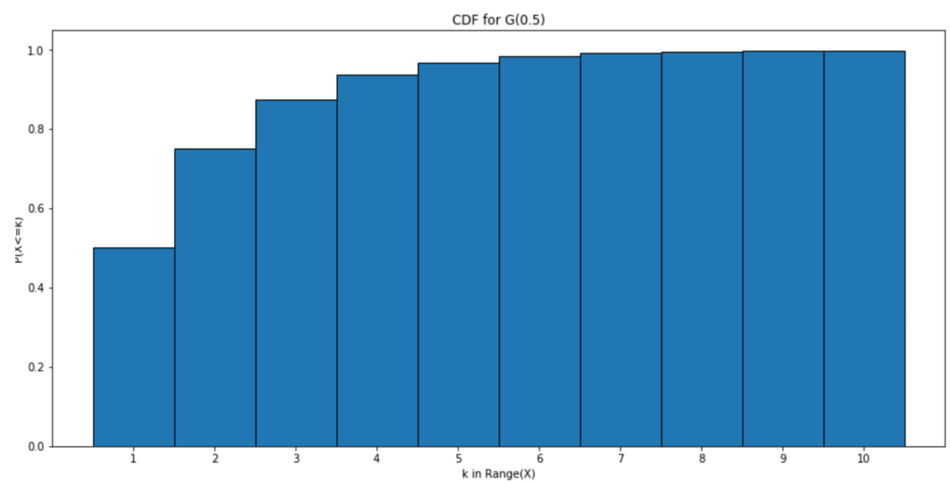
$$R_X = \{ 1, 2, 3, \ldots \}$$

$$f_X = \{ p, (1-p)p, (1-p)^2 p, (1-p)^3 p, \ldots \}$$

$$f_X(k) = (1-p)^{k-1} p$$

$$\begin{aligned} P(X > k) &= (1-p)^k p + (1-p)^{k+1} p + (1-p)^{k+2} p + \ldots \\ &= (1-p)^k \left(p + (1-p)p + (1-p)^2 p + (1-p)^3 p + \ldots \right) \\ &= (1-p)^k \end{aligned}$$

$$P(X \leq k) = F_X(k) = 1.0 - (1-p)^k$$



Geometric Distribution

Example

From an ordinary deck of 52 cards we draw cards at random and *with replacement*, and successively until an ace is drawn. What is the probability that at least 10 draws are needed?

Solution: The probability of an ace is $4/52 = 1/13$. $P(X>9) = (1-1/13)^9 = 0.4866$