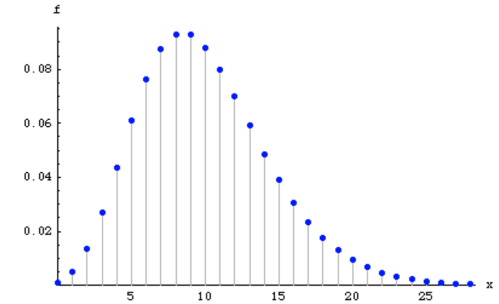
ELEC 2600H: Probability and Random Processes in Engineering



Part II: Single Random Variables

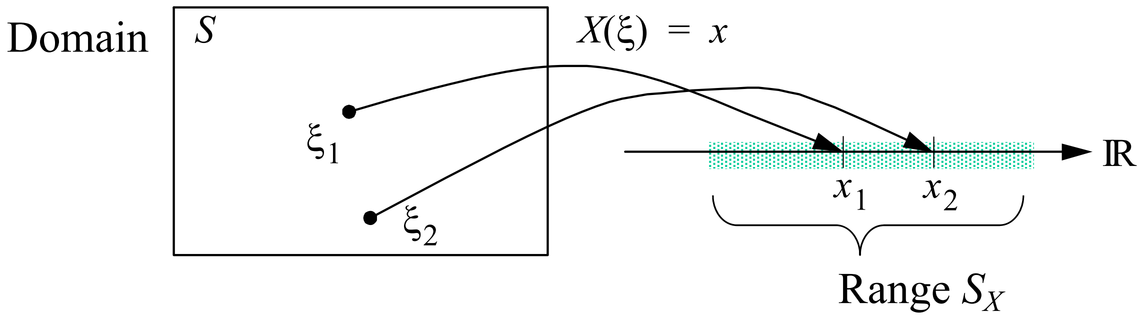
- **Lecture 5: Discrete Random Variables**
- Lecture 6: Expected Value and Moments; Important Discrete Random Variables
- Lecture 7: Continuous Random Variables
- Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs
- Lecture 9: Function of a Random Variable
- Out-of-Class Reading: Conditional PMF and Expectation

Elec 2600H: Lecture 5

- **Random Variables**
  - Equivalent events
- Discrete Random Variables
  - Probability mass function

Random Variables

- **A *random variable* X is a function that assigns a number to every *outcome* of an experiment.**
- The function is fixed and deterministic. All randomness in the observed value is due to the underlying experiment



- Underlying sample space S is called the **domain** of the random variable.
- Set of all possible values of X is called the **range** of the random variable, S_X.

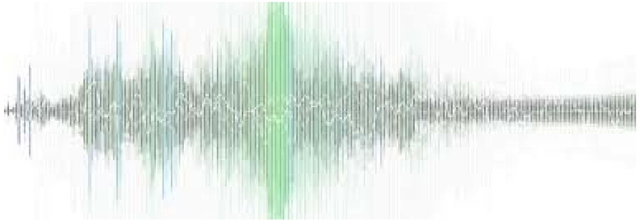
Examples

- Toss a fair die. Take $X = 10i$, where i is the number of dots on the die face

- Transmit either a 5V pulse or a -5V pulse over a transmission line. Take $X =$ "output voltage sampled at the other end."

- Pick a person randomly in the world. Take $X =$ "the height of that person"

- Record a speech waveform by measuring the voltage at the output of a microphone amplifier. Take $X =$ voltage at time 0.

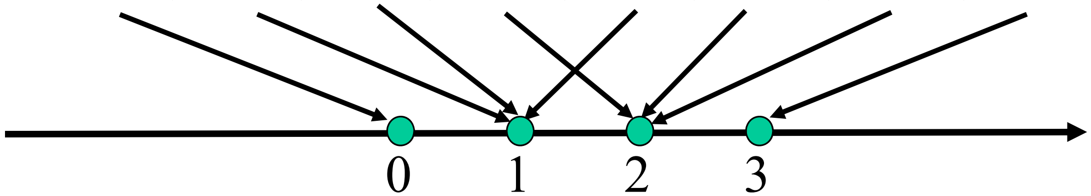


Example 3.1: Coin Tosses

Suppose we toss a fair coin three times. Let X be the number of times heads appears.

Sample space: TTT, TTH, THT, THH, HTT, HTH, HHT, HHH

Real line:

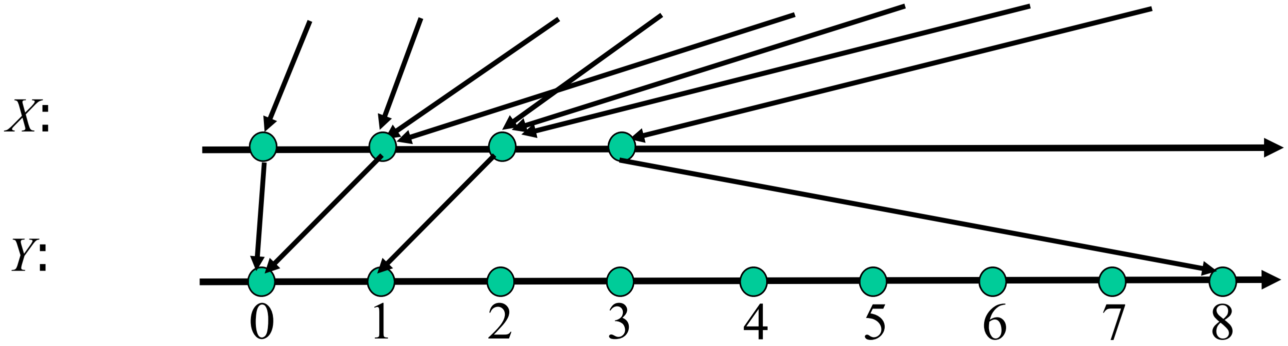


ξ	TTT	TTH	THT	THH	HTT	HTH	HHT	HHH
$X(\xi)$	0	1	1	2	1	2	2	3

Example 3.2: A Betting Game

Let X **be the number of heads** in three tosses of a fair coin. The player receives \$1 if $X = 2$, \$8 if $X = 3$ and nothing otherwise. Let Y **be the reward** to the player.

Sample space: TTT, TTH, THT, THH, HTT, HTH, HHT, HHH



ξ	TTT	TTH	THT	THH	HTT	HTH	HHT	HHH
$X(\xi)$	0	1	1	2	1	2	2	3
$Y(\xi)$	0	0	0	1	0	1	1	8

Elec 2600H: Lecture 5

- **Random Variables**
  - **Equivalent events**

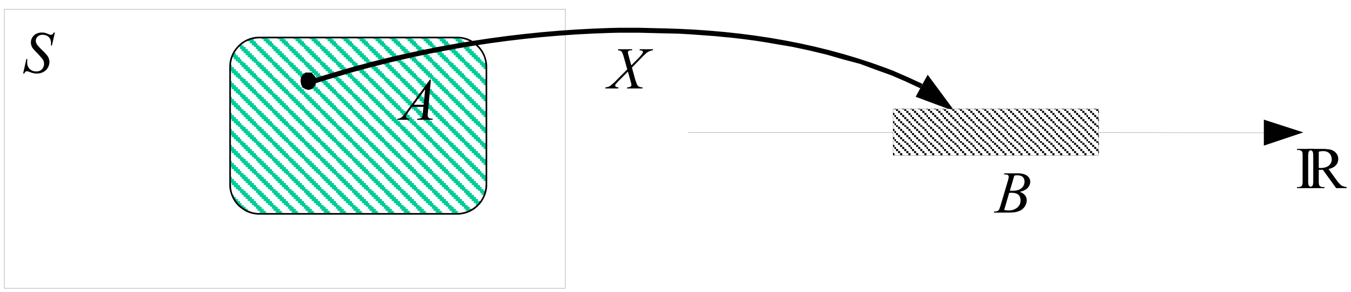
- Discrete Random Variables
  - Probability mass function

Equivalent Event

- Events on the real line inherit probabilities from the original experiment.
- The probability of an event $B \subset \mathbb{R}$ is the probability of the equivalent event:

This is the set of **all** outcomes that map to B. These two events are equivalent because if B occurs, then A must also occur and vice versa.

$$\boxed{A = \{\zeta : X(\zeta) \in B\}}$$



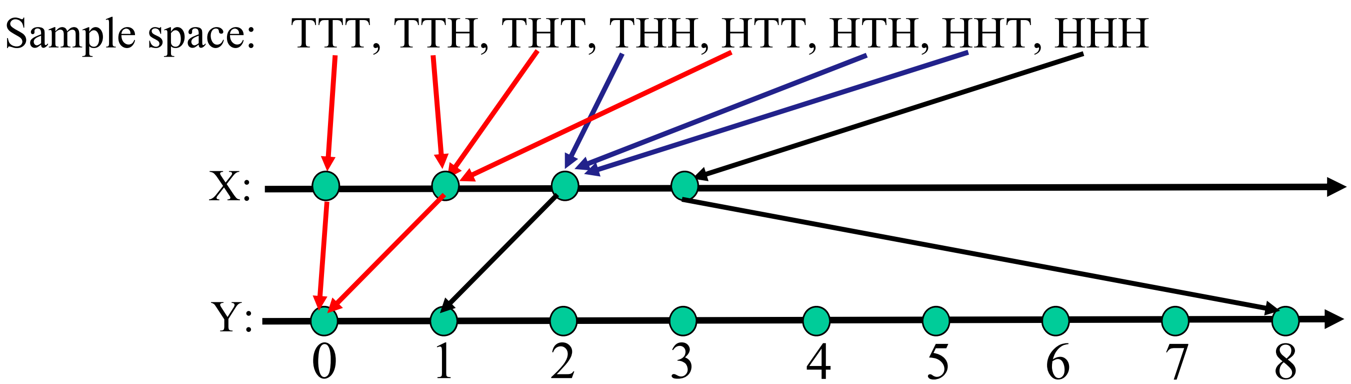
Example 3.3

Let X be the number of heads in three tosses of a fair coin.

The player receives \$1 if $X = 2$, \$8 if $X = 3$ and nothing otherwise.

Let Y be the reward to the player.

Find the probability of the events **$\{X = 2\}$** and **{the player gets nothing}**



Event	Equivalent Event	Probability
$\{X = 2\}$	{THH,HTH,HHT}	$3/8$
{player gets nothing}	{TTT,TTH,THT,HTT}	$4/8 = 1/2$

Elec 2600H: Lecture 5

- Random Variables
  - Equivalent events

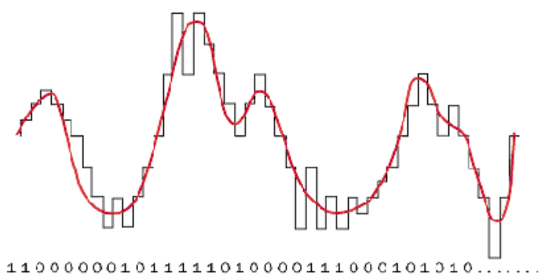
- **Discrete Random Variables**
  - Probability mass function

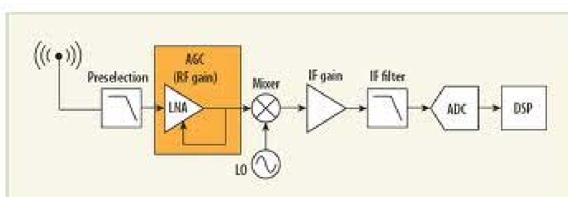
Discrete Random Variables

- A **discrete random variable** assumes values from a countable set $S_X = \{x_1, x_2, x_3, \ldots\}$
- A discrete random variable is **finite** if its range is finite, i.e., $S_X = \{x_1, x_2, x_3, \ldots, x_n\}$
- Note that any random variable defined on an experiment with a finite or countable number of outcomes (e.g. coin flips) must be discrete.

Discrete Random Variables

- However, discrete random variables may also arise from experiments with an uncountably infinite number of outcomes.
- Example: [**Analog to digital conversion (ADC)**]:
  - o The original voltage is defined on the real line, but the output of the ADC assumes only a finite number of values (e.g. 256 for 8 bit resolution)





5. This block diagram shows the typical architecture of an RF communications receiver.

Specifying Probabilities of Discrete Random Variables

- Often, we wish to compute probabilities of events directly, without finding equivalent events in the original sample space
  - It's easier.
  - The underlying probability space may be too complex to analyze or describe.
- The probabilities of any event based on a discrete random variable can be computed in various ways:
  1. **from the probability mass function (pmf)**
  2. **from the cumulative distribution function (cdf)**
  3. **from the characteristic function**
- These are *minimal* in the sense they contain enough information to compute any probability, but no extra information.
- We usually use the **probability mass function**…

Elec 2600H: Lecture 5

- Random Variables
  - Equivalent events

- **Discrete Random Variables**
  - **Probability mass function**

Probability Mass Function (pmf)

- Consider a discrete random variable X that assumes values from a finite or countable set

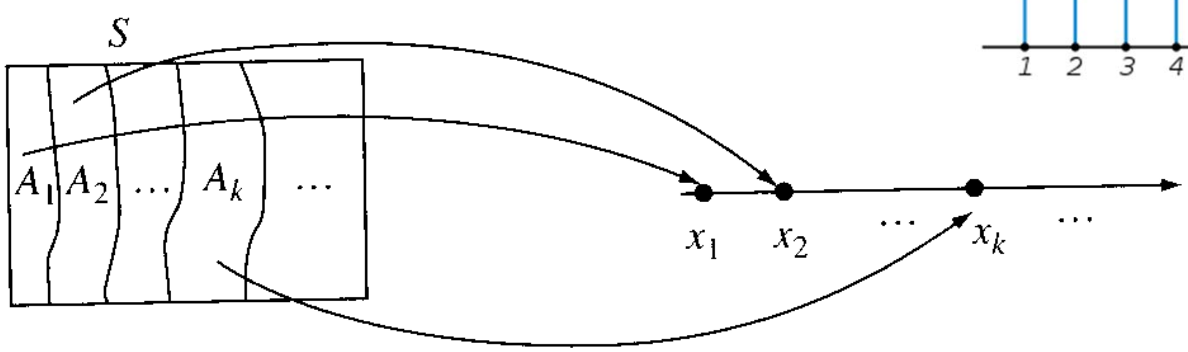
$$S_X = \{x_1, x_2, x_3, \ldots\}$$

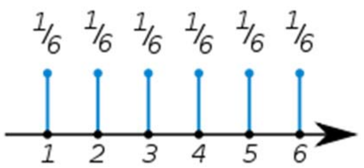
- The ***probability mass function (pmf)*** of X is defined as:

$$\boxed{p_X(x) = P\left[X = x\right] = P\left[\{\zeta : X(\zeta) = x\}\right]}$$

- Note that $p_X(x)$ is zero almost everywhere. It can be nonzero only at the values $x_1, x_2, x_3, \ldots$. At these values,

$$\boxed{p_X(x_k) = P\left[A_k\right] \quad \text{where } A_k = \{\zeta : X(\zeta) = x_k\}}$$



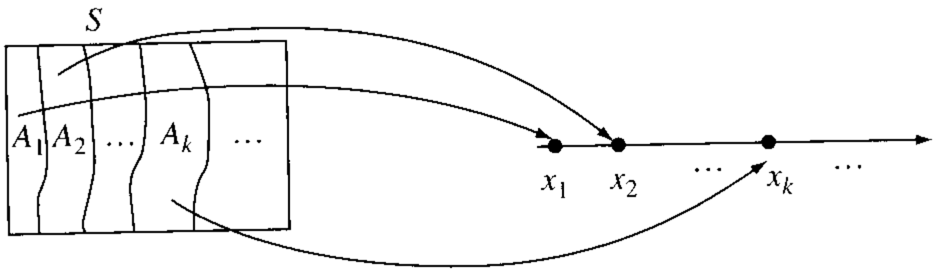


Properties of Probability Mass Functions

- $\boxed{p_X(x) \geq 0}$ (since probabilities are **non-negative**)

- $\boxed{\sum_{x \in S_X} p_X(x) = \sum_{\text{all } k} p_X(x_k) = \sum_{\text{all } k} P[A_k] = 1}$

  since the A_k partition the sample space.



- $\boxed{P[X \text{ in } B] = \sum_{x \in B} p_X(x) \text{ where } B \subset S_X}$

  since $\{X \text{ in } B\} = \bigcup_{k : x_k \in B} A_k$

Example 3.5: Coin Tosses

Find the pmf of X, the number of heads in 3 tosses of a fair coin.

Solution:

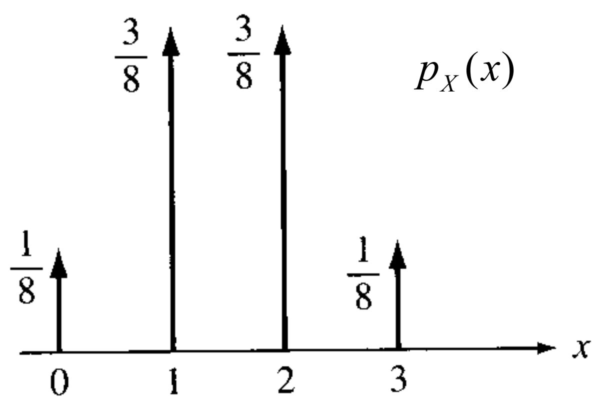
$$p_X(0) = P[X = 0] = P[\{\text{TTT}\}] = \frac{1}{8}$$

$$p_X(1) = P[X = 1] = P[\{\text{HTT,THT,TTH}\}] = \frac{3}{8}$$

$$p_X(2) = P[X = 2] = P[\{\text{HHT,HTH,THH}\}] = \frac{3}{8}$$

$$p_X(3) = P[X = 3] = P[\{\text{HHH}\}] = \frac{1}{8}$$

Note that $p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1$



<u>Example 3.6: A Betting Game</u>

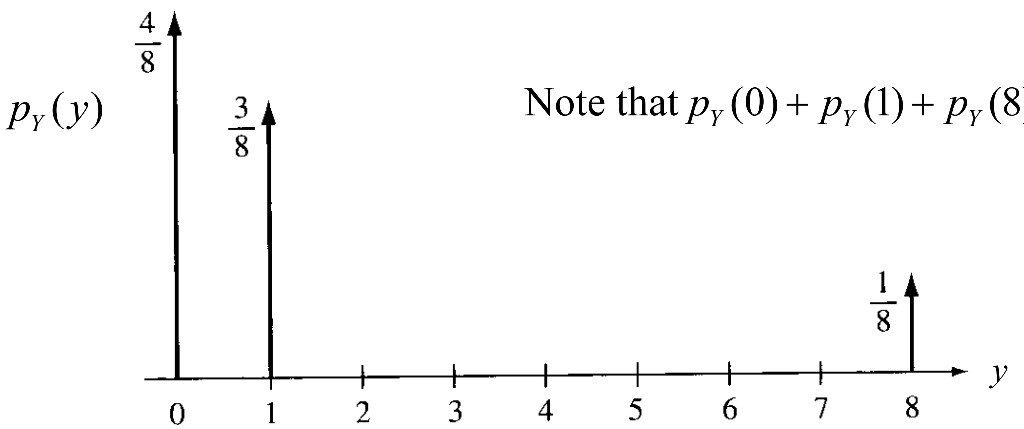
Let X be the number of heads in three tosses of a fair coin.
The player receives \$1 if $X = 2$, \$8 if $X = 3$ and nothing otherwise.
Find the pmf of Y, the reward to the player.

Solution:

$$p_Y(0) = P[\{X = 0\} \cup \{X = 1\}] = P[\{\text{TTT},\text{HTT},\text{THT},\text{TTH}\}] = \tfrac{4}{8} = \tfrac{1}{2}$$

$$p_Y(1) = P[X = 2] = P[\{\text{HHT},\text{HTH},\text{THH}\}] = \tfrac{3}{8}$$

$$p_X(8) = P[X = 3] = P[\{\text{HHH}\}] = \tfrac{1}{8}$$



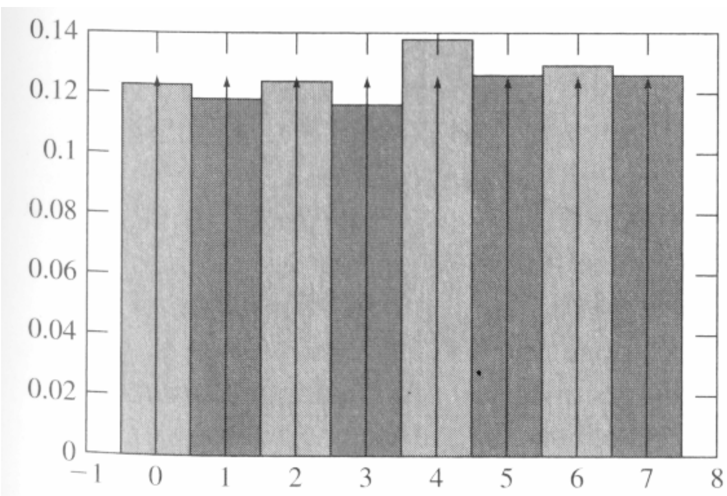
Note that $p_Y(0) + p_Y(1) + p_Y(8) = 1$

Example 3.7: Discrete Uniform RV

Suppose a *random number generator* produces an integer number X at random (with equal probability) from the set $S_X = \{0,1,2,\ldots,M-1\}$. Find the pmf of X.

Solution: Since there are M outcomes and they are equally likely $p_X(k) = \frac{1}{M}$ where $k \in S_X$

Intuitively, if we pick many random numbers, the relative frequency of each number is about the same.



$M = 8$

Bars show relative frequency computed over 1000 trials

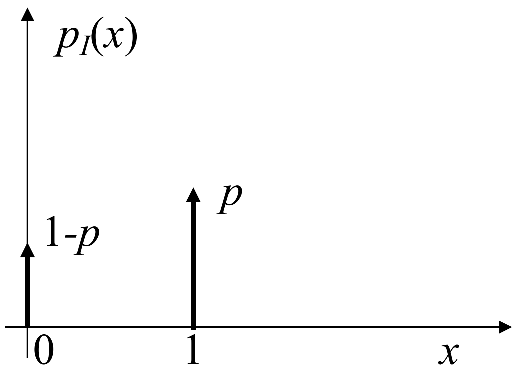
Arrows show pmf

x

Example 3.8: Bernoulli Random Variable

- Let A be an event of interest in a random experiment (i.e. heads on a coin flip, a device is working)
- We say that a "success" occurs if A occurs.
- The **Bernoulli RV** is the **indicator function** for A, defined by $I_A(\zeta) = \begin{cases} 0 & \text{if } \zeta \notin A \text{ ("failure")} \\ 1 & \text{if } \zeta \in A \text{ ("success")} \end{cases}$
- I_A is a discrete RV with pmf

$$p_I(0) = 1 - p \qquad \text{where } p = P[A]$$
$$p_I(1) = p$$



Example 3.9: Message Transmissions

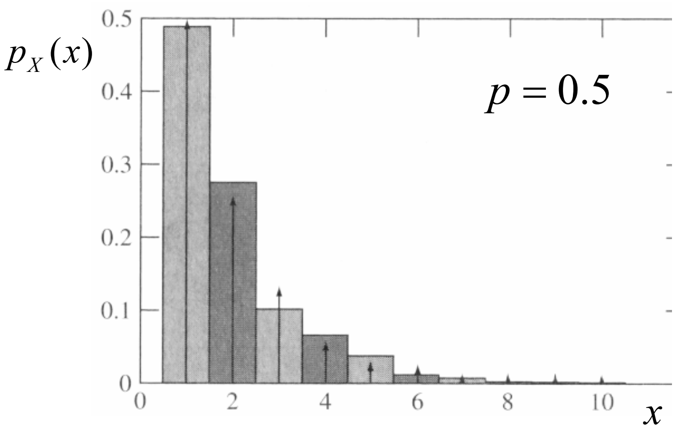
Suppose a message is transmitted until it arrives correctly at its destination. Assume that errors occur independently and with probability $1\text{-}p$ each time the message is transmitted (i.e. the message is received correctly with probability p)

Let X be the number of times it needs to be transmitted. Find the pmf of X.

Solution:

X assumes values in $S_X = \{1, 2, 3, \ldots\}$. The event $\{X=k\}$ occurs if there are $k\text{-}1$ erroneous transmissions followed by a success.

$$p_X(k) = P[X = k] = (1-p)^{k-1}\, p \qquad \text{for } k = 1, 2, \ldots., \quad \textbf{(Geometric RV)}$$



Bars show relative frequency computed over 1000 trials

Arrows show pmf

Example 3.9: Message Transmissions

For the experiment described on the previous slide, find the probability that X is even.

Solution:

$$P[X \text{ is even}] = \sum_{j \text{ even}} p_X(j) = \sum_{k=1}^{\infty} p_X(2k) = \sum_{k=1}^{\infty} (1-p)^{2k-1} p = p(1-p) \sum_{k=1}^{\infty} \left((1-p)^2\right)^{k-1}$$

$$= \frac{p(1-p)}{1-(1-p)^2} = \frac{p(1-p)}{1-(1-2p+p^2)} = \frac{p(1-p)}{2p-p^2} = \frac{1-p}{2-p}$$

$$\boxed{\sum_{k=1}^{\infty} x^{k-1} = \frac{1}{1-x}}$$

Example 3.10: Transmission Errors

A binary communication channel introduces a bit error in a transmission with probability p. Let X be the number of errors in n independent transmissions.

Find the pmf of X and the probability of one or fewer errors.

Solution:

X takes on values in $S_X = \{0, 1, 2, 3, \ldots, n\}$

The probability of k errors in n transmissions is the probability of any pattern of k errors and n-k correct transmissions.

$$p_X(k) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, 2, \ldots., n \qquad \textbf{(Binomial RV)}$$

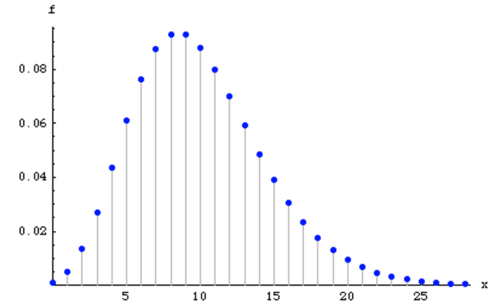
The probability of one or fewer errors is

$$P[X \le 1] = p_X(0) + p_X(1) = \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1}$$

$$= (1-p)^n + np(1-p)^{n-1}$$

ELEC 2600H: Probability and Random Processes in Engineering

Part II: Single Random Variables

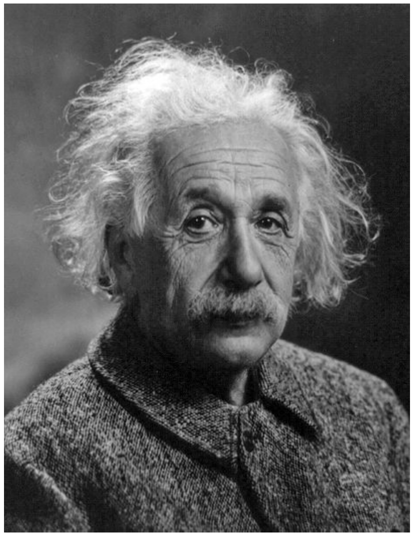


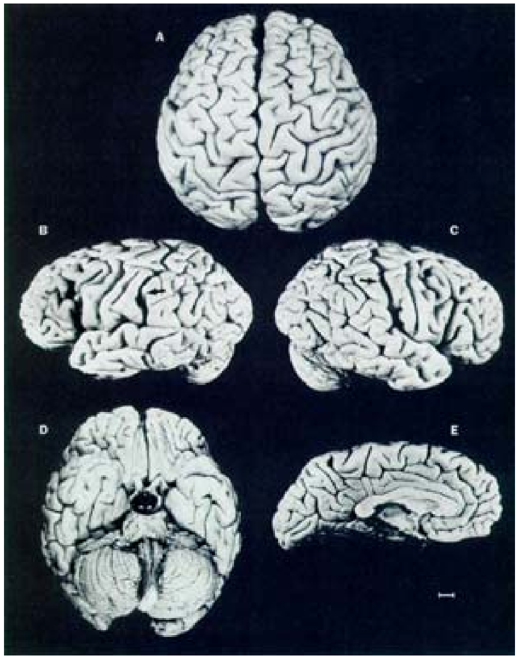
- Lecture 5: Discrete Random Variables
- **Lecture 6: Expected Value and Moments; Important Discrete Random Variables**
- Lecture 7: Continuous Random Variables
- Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs
- Lecture 9: Function of a Random Variable
- Out-of-Class Reading: Conditional PMF and Expectation

Elec 2600H: Lecture 6

- **Expectation of a random variable**
- Expected value of a function of a random variable
- Variance of a random variable
- Moments of a random variable
- Important discrete random variables
  - Summary of variables you know:
    - **Bernoulli**
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Interesting Fact… Einstein's Brain





Einstein's brain weighs less than average, at 1230 grams!!

Definition of the Expected Value

- Expected value of a discrete random variable defined by

$$m_X = E[X] = \sum_{x \in S_X} x p_X(x) = \sum_k x_k p_X(x_k)$$

  - The expected value is defined if the sum above converges absolutely:

$$\sum_k |x_k| p_X(x_k) < \infty$$

  - If this sum does not converge, then the expected value does not exist

- If we view $p_X(x)$ as the distribution of mass on the points $x_1, x_2, x_3, \ldots$ then the expected value is the center of mass of the distribution.

Example 3.11: Mean of Bernoulli RV

- The **Bernoulli random variable** assumes two values: 0 and 1 with probabilities

$$p_X(0) = (1-p) \quad p_X(1) = p$$

- Using the definition

$$m_X = E[X] = \sum_{x \in S_X} x p_X(x) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p$$

- Note that we should ***NOT*** interpret the "expected value" as the ***value we expect to see*** on any single trial, since in the example above, the expected value is not *even one of the possible values* if $0 < p < 1$.

Example 3.12

Let X be the number of heads in three tosses of a fair coin. Find $\mathrm{E}[X]$.

Solution

The sample space is $S_X = \{0,1,2,3\}$
The probability mass function is given by

$$p_X(0) = \tfrac{1}{8} \quad p_X(1) = \tfrac{3}{8} \quad p_X(2) = \tfrac{3}{8} \quad p_X(3) = \tfrac{1}{8}$$

By the definition

$$\begin{aligned} E[X] &= \sum_{x \in S_X} x p_X(x) \\ &= 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) + 3 \cdot p_X(3) \\ &= 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \frac{(0+3+6+3)}{8} = 1.5 \end{aligned}$$

Example 3.13: Uniform Discrete RV

Let X be a **discrete uniform random variable** assuming values in the set $S_X = \{0, 1, 2, ..., M\text{-}1\}$. Find $\mathrm{E}[X]$.

Solution

Uniform pmf is given by:

$$p_X(k) = \tfrac{1}{M} \quad \text{for } k = 0, 1, \ldots, M-1$$

Substituting into the definition,

$$E[X] = \sum_{x \in S_X} x \cdot p_X(x) = \sum_{k=0}^{M-1} k \cdot p_X(k)$$

$$= \sum_{k=0}^{M-1} k \cdot \frac{1}{M} = \frac{1}{M}\left(\sum_{k=0}^{M-1} k\right)$$

$$= \frac{1}{M}\left(\frac{(M-1)M}{2}\right) = \frac{M-1}{2}$$

Example:
If $M = 3$,
$S_X = \{0, 1, 2\}$
and $E[X] = 1$

Relative Frequency Interpretation

- $E[X]$ corresponds to the "average of X" in a large number of observations of X.
- Suppose that we perform an experiment n times.
  - Let $x(j)$ be the value of the RV in the j th experiment.
  - Let $N_k(n)$ be the number of times that value x_k is observed.
- We can compute the average either by summing the observations in order, or by counting the number of times each outcome occurs.

$$\langle X\rangle_n = \frac{x(1)+x(2)+\ldots+x(n)}{n} = \frac{x_1 N_1(n) + x_2 N_2(n) + \cdots + x_k N_n(n) + \cdots}{n}$$

$$= x_1 \frac{N_1(n)}{n} + x_2 \frac{N_2(n)}{n} + \cdots + x_k \frac{N_n(n)}{n} + \cdots$$

$$= x_1 f_1(n) + x_2 f_k(n) + \cdots + x_k f_k(n) + \cdots = \sum_k x_k f_k(n)$$

Relative Frequency Interpretation (cont.)

- From the previous page, the average value of X in n trials is

$$\boxed{\langle X \rangle_n = \sum_k x_k f_k(n)}$$

- As the number of trials, n, increases, $\lim_{n\to\infty} f_k(n) = p_X(x_k) \text{ for all } k$

- This implies that

$$\boxed{\langle X \rangle_n = \sum_k x_k f_k(n) \xrightarrow[n\to\infty]{} \sum_k x_k p_X(x_k) = E[X]}$$

Example 3.14

X	0	1	2	3
p_X	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
Y	0	0	1	8

Suppose a player flips a fair coin 3 times

Let X be the number of heads that appears

Let Y be the reward, which is \$1 for 2 heads, \$8 for 3 heads, and nothing otherwise.

Find the expected reward, $E[Y]$.

Solution

From the definition,
$$E[Y] = \sum_{y \in S_Y} y \times p_Y(y)$$

Thus, we must find the **pmf of Y**:
$$p_Y(0) = P[X = 0 \text{ or } X = 1] = \tfrac{1}{8} + \tfrac{3}{8} = \tfrac{4}{8}$$
$$p_Y(1) = P[X = 2] = \tfrac{3}{8}$$
$$p_Y(8) = P[X = 3] = \tfrac{1}{8}$$

Substituting, we obtain
$$E[Y] = \$0 \times p_Y(0) + \$1 \times p_Y(1) + \$8 \times p_Y(8)$$
$$= \$0 \times \frac{4}{8} + \$1 \times \frac{3}{8} + \$8 \times \frac{1}{8} = \$\frac{11}{8} = \underline{\$1.375}$$

You should not pay more than \$1.375 to play this game!!!

Example 3.15: Mean of Geometric RV

Suppose that X, the number of tries until a successful message transmission, has a **geometric distribution** with parameter p:

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

$$p_X(k) = p(1-p)^{k-1} \quad \text{for } k \in S_X$$

Find the mean of X, $\mathrm{E}[X]$.

Solution

Substituting into the definition,

$$E[X] = \sum_{k=1}^{\infty} k \cdot p_X(k) = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1}$$

$$= p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\sum_{k=0}^{\infty} k(1-p)^{k-1}$$

$$= p\frac{1}{(1-(1-p))^2} = p\frac{1}{p^2} = \frac{1}{p}$$

$$\boxed{\begin{gathered}\text{Since } \sum_{k=0}^{\infty} a^k = \frac{1}{1-a}, \\ \frac{d}{da}\left(\sum_{k=0}^{\infty} a^k\right) = \frac{d}{da}\left(\frac{1}{1-a}\right) \\ \sum_{k=0}^{\infty} k a^{k-1} = \frac{1}{(1-a)^2}\end{gathered}}$$

Intuitively, average number of tries becomes less as p increases!

Example 3.27 Mean of Binomial RV

Suppose that X is a **binomial RV** with parameters n and p. Find $E[X]$.

Solution

Binomial pmf: $p_X(k) = P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0,1,2,\ldots,n$

$$E[X] = \sum_{k=0}^{n} k p_X(k) = \sum_{k=1}^{n} k p_X(k)$$

$$= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{(k-1)} (1-p)^{n-k}$$

$$= np \underbrace{\sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} p^j (1-p)^{n-1-j}}_{=1}$$ make a change of variables $j = k-1$, $k = j+1$

$= np$

=1 since sum of pmf of binomial RV with parameters n-1 and p

number of trials × probability of success on each trial

Elec 2600H: Lecture 6

- Expectation of a random variable
- **Expected value of a function of a random variable**
- Variance of a random variable
- Moments of a random variable
- Important discrete random variables
  - Summary of variables you know:
    - **Bernoulli**
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Expected Value of Functions of a RV

- Let X be a random variable and let $\mathbf{Z = g(X)}$.
  - Let $S_X = \{x_1, x_2, \ldots, x_k, ..\}$ be the set of possible values of X.
  - Let $S_Z = \{z_1, z_2, \ldots, z_k, \ldots\}$ be the set of possible values of Z.
- By definition, we can find the expected value of Z by $E[Z] = \sum_k z_k p_Z(z_k)$
- To use this formula, given the pmf of X, we must first find the pmf of Z according to the formula $$p_Z(z_j) = \sum_{x_k : g(x_k) = z_j} p_X(x_k)$$
- But, there is an easier way, that **uses the pmf of X to find $E[Z]$ <u>directly</u>**:

$$\boxed{E[Z] = E[g(X)] = \sum_k g(x_k) p_X(x_k)}$$

Proof:

$$\sum_k g(x_k) p_X(x_k) = \sum_j z_j \left\{ \sum_{x_k : g(x_k) = z_j} p_X(x_k) \right\} = \sum_j z_j p_Z(z_j) = E[Z]$$

rearrange sum to group all terms x_k
that map to the same value of z_j

Example: Betting Game

X	0	1	2	3
p_X	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
Y	0	0	1	8

- Suppose a player flips a fair coin 3 times
  - Let X be the number of heads that appears
  - Let Y be the reward, which is \$1 for 2 heads, \$8 for 3 heads, and nothing otherwise.
- Find the expected reward, $\mathrm{E}[Y]$.

<u>Solution:</u>

Using our new formula and letting $g(.)$ be the function mapping X to Y,

$$E[Y] = E[g(X)] = \sum_k g(x_k) p_X(x_k)$$

$$= \$0 \times p_X(0) + \$0 \times p_X(1) + \$1 \times p_X(2) + \$8 \times p_X(3)$$

$$= \$0 \times \frac{1}{8} + \$0 \times \frac{3}{8} + \$1 \times \frac{3}{8} + \$8 \times \frac{1}{8} = \$\frac{11}{8} = \$1.375$$

Example 3.17: Square Law Device

Let X be a noise voltage that assumes values in $S_X = \{-3, -1, 1, 3\}$ with equal probability $(1/4)$. Find $\mathrm{E}[Z]$ where $\boldsymbol{Z = X^2}$.

Solution

[Method 1] Find the pmf of Z and then use the definition:

$$p_Z(z) = \begin{cases} p_X(-1) + p_X(1) = \tfrac{1}{2} & \text{if } z = 1 \\ p_X(-3) + p_X(3) = \tfrac{1}{2} & \text{if } z = 9 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[Z] &= \sum_k z_k p_Z(z_k) \\ &= 1 \times p_Z(1) + 9 \times p_Z(9) \\ &= 1 \times \tfrac{1}{2} + 9 \times \tfrac{1}{2} = \frac{10}{2} = \underline{5} \end{aligned}$$

[Method 2] Use the direct formula:

$$\begin{aligned} E[Z] = E[X^2] &= \sum_k x_k{}^2 p_X(x_k) \\ &= (-3)^2 \times p_X(-3) + (-1)^2 \times p_X(-1) + (1)^2 \times p_X(1) + (3)^2 \times p_X(3) \\ &= 9 \times \tfrac{1}{4} + 1 \times \tfrac{1}{4} + 1 \times \tfrac{1}{4} + 9 \times \tfrac{1}{4} = \frac{20}{4} = \underline{5} \end{aligned}$$

Useful Results

- **Scaling a RV by a constant** a scales the mean by a.

$$\boxed{E[aX] = \sum_k a x_k p_X(x_k) = a \sum_k x_k p_X(x_k) = aE[X]}$$

- **The mean of a constant** c is the constant.

$$\boxed{E[c] = \sum_k c p_X(x_k) = c \sum_k p_X(x_k) = c} \quad \text{since } \sum_k p_X(x_k) = 1$$

- **Shifting a RV by a constant** c shifts the mean by c:

$$\boxed{E[X+c] = \sum_k (x_k + c) p_X(x_k) = \sum_k x_k p_X(x_k) + c \sum_k p_X(x_k) = E[X] + c}$$

- In general, for functions $g(\cdot)$ and $h(\cdot)$ and constants a, b and c

$$\boxed{E[ag(X) + bh(X) + c] = aE[g(X)] + bE[h(X)] + c}$$

However:

$$E[ag(X) + bh(X) + c] \neq ag(E[X]) + bh(E[X]) + c$$

Example 3.18: Noise Voltage

Let X be a noise voltage that assumes values in $S_X = \{-3, -1, 1, 3\}$ with equal probability $(1/4)$.
Find $E[Z]$ where $Z = (2X+10)^2$.

Solution

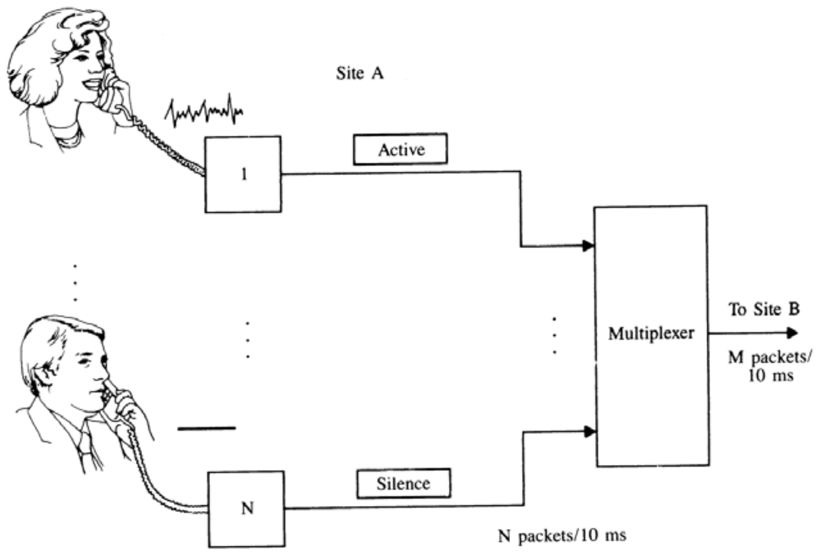
Exploit linearity as much as possible!

$$\begin{aligned} E[Z] &= E[(2X+10)^2] \\ &= E[4X^2 + 40X + 100] \\ &= 4E[X^2] + 40E[X] + 100 \\ &= 4(5) + 40(0) + 100 = 120 \end{aligned}$$

Note that you **CANNOT** simply substitute the expected value of X into the expression for Z: $E[Z] \neq (2E[X]+10)^2 = (2\times 0+10)^2 = 100$

Example 3.19: Packet Voice Transmission

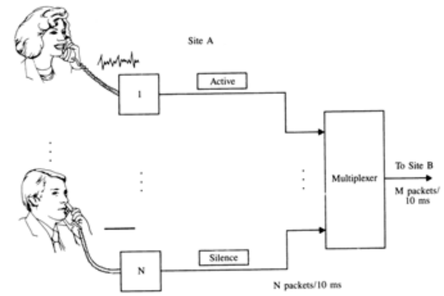
- Suppose a communication system needs to transmit $n = 48$ simultaneous conversations using "packets" corresponding to 10ms of speech.
- Suppose each person speaks independently only $1/3$ of the time.
- Suppose that the packet multiplexer transmits up to $\boldsymbol{m = 20}$ packets every 10ms. Excess packets are discarded.
- Let Z be the number of packets discarded. Find $\mathrm{E}[Z]$.



Example 3.19 Solution

- Since there are $n = 48$ speakers, each speaking independently and with probability $1/3$, the number of **active speech packets**, X, is a **binomial random variable** with parameters n=48 and p=1/3:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0,1,2,\ldots,n$$



- The number of **discarded packets**, Z, is a function of X:

$$Z = (X-20)^+ = \begin{cases} 0 & \text{if } X \le 20 \\ X-20 & \text{if } X > 20 \end{cases}$$

- Thus,

$$E[Z] = \sum_{k=0}^{48} (k-20)^+ \cdot p_X(k)$$

$$= \sum_{k=21}^{48} (k-20) \cdot p_X(k)$$

$$= \sum_{k=21}^{48} (k-20) \cdot \binom{48}{k} \left(\tfrac{1}{3}\right)^k \left(\tfrac{2}{3}\right)^{n-k} \approx 0.182$$

Much smaller than the expected number of active packets:
E[X] = np = 16

Elec 2600H: Lecture 6

- Expectation of a random variable
- Expected value of a function of a random variable
- **Variance of a random variable**
- Moments of a random variable
- Important discrete random variables
  - Summary of variables you know:
    - **Bernoulli**
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Variance

- The variance of a random variable measures how much the values of the random variable generally differ from the mean.
- Define D to be the difference between the value of the random variable and its mean:

$$D = X - E[X]$$

  Note that this can be positive or negative.
- The *variance* of a RV, $\mathrm{VAR}[X]$ (often denoted by s^2), is defined as the expected value of the squared difference:

$$\boxed{\mathrm{VAR}[\mathrm{X}] = E\left[D^2\right] = E[(X - E[X])^2]}$$

- The *standard deviation* of a RV, $\mathrm{STD}[X]$ (often denoted by s), is defined as the square root of the variance.

$$\boxed{\mathrm{STD}[\mathrm{X}] = \sqrt{\mathrm{VAR}[\mathrm{X}]}}$$

  The advantage of the standard deviation is that its units are the same as that of X. For example, the variance of height would be in units of meters^2, but the standard deviation would be in units of meters

Properties of the Variance

- $\boxed{\mathrm{VAR}[X] = E[X^2] - (E[X])^2}$ **(VERY useful formula!)**

Proof: $\mathrm{VAR}[X] = E\left[(X - E[X])^2\right] = E\left[X^2 - 2X \cdot E[X] + E[X]^2\right]$

$= E\left[X^2\right] - 2E[X] \cdot E[X] + E[X]^2 = E\left[X^2\right] - E[X]^2$

- $\boxed{\mathrm{VAR}[c] = 0}$

Proof: $\mathrm{VAR}[c] = E\left[c^2\right] - E[c]^2 = c^2 - c^2 = 0$

Properties of the Variance (cont.)

- VAR[X+c] = VAR[X]

Proof:

$$\begin{aligned}\mathrm{VAR}[X+c] &= E\left[(X+c)^2\right]-E[(X+c)]^2 \\ &= E\left[(X^2+2cX+c^2)\right]-(E[X]+c)^2 \\ &= E\left[X^2\right]+2cE[X]+c^2-\left(E[X]^2+2cE[X]+c^2\right) \\ &= E\left[X^2\right]-E[X]^2=\mathrm{VAR}[X]\end{aligned}$$

- VAR[cX] = c^2VAR[X]

Proof:

$$\begin{aligned}\mathrm{VAR}[cX] &= E\left[c^2X^2\right]-E[cX]^2=c^2E\left[X^2\right]-(cE[X])^2 \\ &= c^2E\left[X^2\right]-c^2E[X]^2=c^2\left(E\left[X^2\right]-E[X]^2\right)=c^2\mathrm{VAR}[X]\end{aligned}$$

Elec 2600H: Lecture 6

- Expectation of a random variable
- Expected value of a function of a random variable
- Variance of a random variable
- **Moments of a random variable**
- Important discrete random variables
  - Summary of variables you know:
    - **Bernoulli**
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Moments

- The mean and variance are examples of the ***moments* of a random variable.**
- The moments are numbers (constants) that summarize different information about the pmf. Under certain conditions, knowing all the moments is equivalent to knowing the pmf.

- The n^{th} moment of a random variable, X, is defined as

$$\boxed{E[X^n] = \sum_k x_k^n p_X(x_k)}$$

- The n^{th} central moment of a random variable, X, is defined as

$$\boxed{E\left[(X - E[X])^n\right] = \sum_k \left(x_k - E[X]\right)^n p_X(x_k)}$$

- Important moments:
  - 1st moment, $E[X]$ (the mean)
  - 2nd central moment, $E[(X\text{-}E[X])^2]$ (the variance)

Example 3.20 Three Coin Tosses

Let X be the number of heads in three tosses of a fair coin. Find $\text{Var}[X]$.

Solution

The sample space is $S_X = \{0,1,2,3\}$

The probability mass function is given by $p_X(0) = \frac{1}{8} \quad p_X(1) = \frac{3}{8}$

$p_X(2) = \frac{3}{8} \quad p_X(2) = \frac{1}{8}$

We have already seen that $E[X] = 1.5$

To find the variance, we use the formula

$$\begin{aligned}\text{Var}[X] &= E[X^2] - E[X]^2 \\ &= 3 - (1.5)^2 = 0.75\end{aligned}$$

where $E[X^2] = \sum_k x_k{}^2 p_X(x_k)$

$$= 0^2 \cdot \tfrac{1}{8} + 1^2 \cdot \tfrac{3}{8} + 2^2 \cdot \tfrac{3}{8} + 3^2 \cdot \tfrac{1}{8} = \frac{3+12+9}{8} = \frac{24}{8} = 3$$

Example 3.21: Variance of Bernoulli RV

- The Bernoulli random variable assumes two values: 0 and 1 with probabilities

$$p_X(0) = (1 - p) \quad p_X(1) = p$$

- We have already seen that

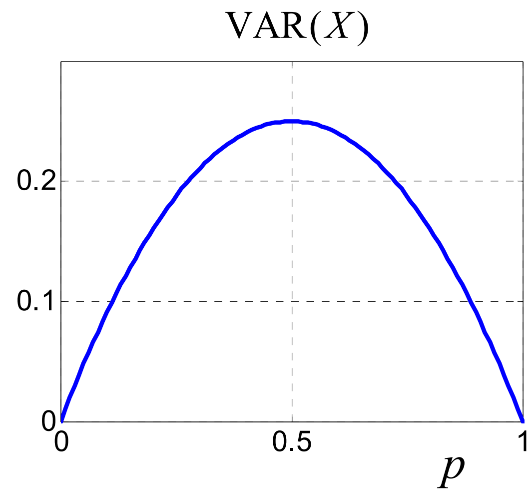
$$E[X] = p$$

- To find the variance, we use the formula

$$\begin{aligned}\mathrm{Var}[X] &= E[X^2] - E[X]^2 \\ &= p - p^2 = p(1-p)\end{aligned}$$

where

$$\begin{aligned}E[X^2] &= \sum_k x_k{}^2 p_X(x_k) \\ &= 0^2 \cdot (1-p) + 1^2 \cdot p \\ &= p\end{aligned}$$



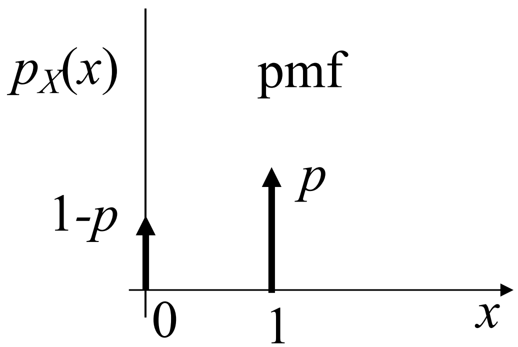
Elec 2600H: Lecture 6

- Expectation of a random variable
- Expected value of a function of a random variable
- Variance of a random variable
- Moments of a random variable
- **Important discrete random variables**
  - **Summary of variables you know:**
    - Bernoulli
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Bernoulli Random Variable

- The Bernoulli Random Variable is simply a random variable that assumes either value 0 or 1 with probabilities $(1\text{-}p)$ and p.

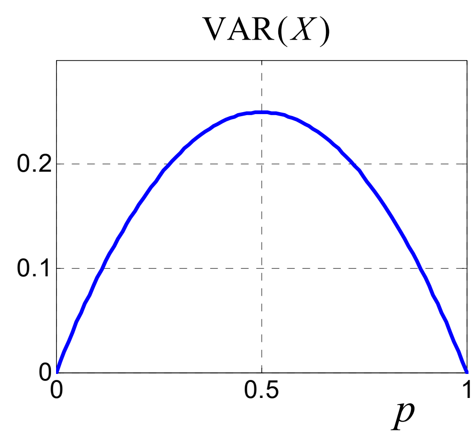
$$p_X(0) = 1 - p \qquad p_X(1) = p$$



- Mean and Variance

$$E[X] = p$$
$$\mathrm{VAR}[X] = p(1-p)$$

- Used to model
  - Single coin toss
  - ***Occurrence*** of an event of interest



Binomial Random Variable

- Suppose a random experiment is repeated n independent times. For each trial, an event A occurs with probability p
- X, **number of times that an event A occurs**, is a binomial RV with

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0,1,\ldots,n$$

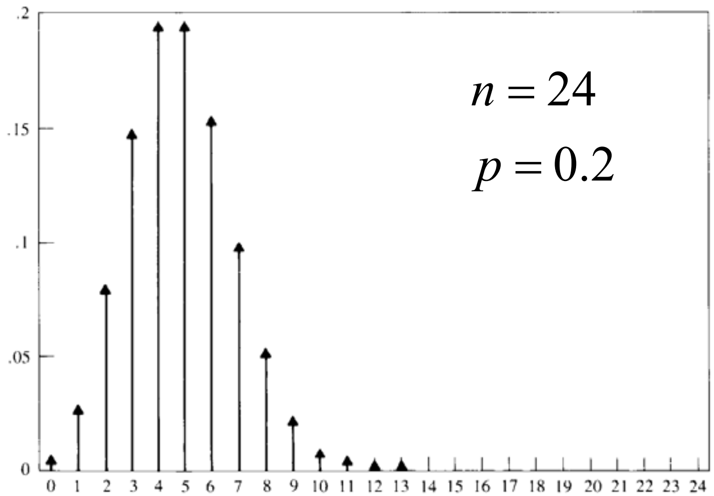
- Mean and Variance

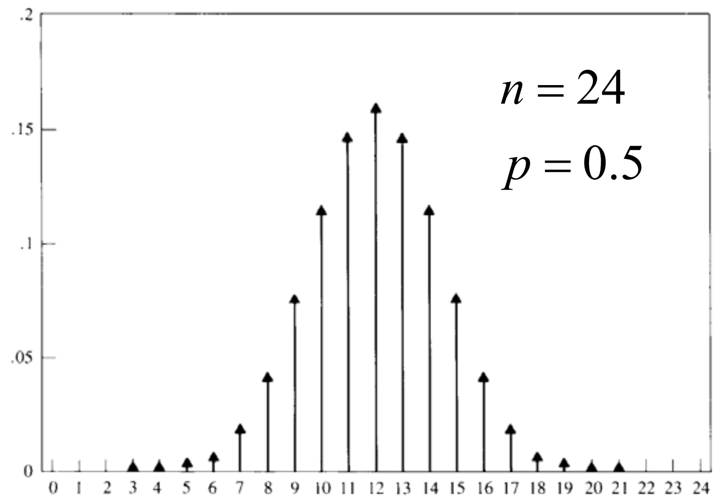
$$E[X] = np$$
$$\mathrm{VAR}[X] = np(1-p)$$

note that these are just n **times the values for the Bernoulli** (more on this later)

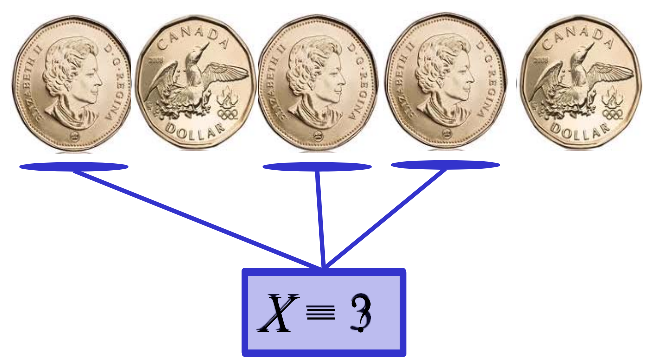
- Applications
  - o Multiple coin flips
  - o Occurrence of a property in individuals of a population (e.g. bit errors in a transmission, defective parts in a batch)

Binomial Random Variable: Example PMFs





Symmetric for $p=0.5$?
$p = 0.8$?



Geometric Random Variable

- Suppose a random experiment is repeated **until** an event A occurs. In each repeat, A occurs independently and with probability p.
- The number of trials until the first success, M, is a *geometric RV*

$$\boxed{p_M(k) = (1-p)^{k-1}\, p \qquad \text{for } k = 1, 2, \ldots.,}$$

- Mean and variance: $\boxed{E[M] = \frac{1}{p}}$ $\boxed{\mathrm{VAR}[M] = \frac{1-p}{p^2}}$

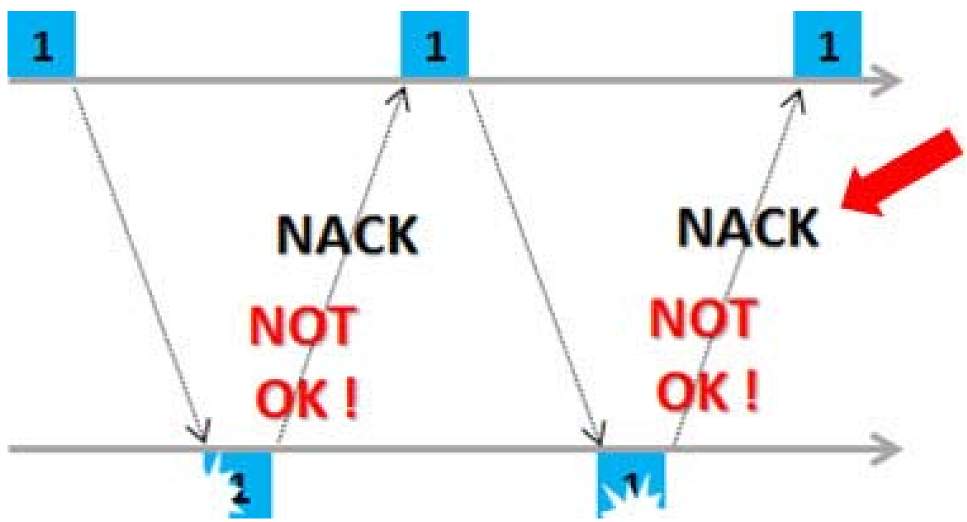
- The number of failures before the first success, $M' = M\text{-}1$ occurs is also called a geometric RV.

$$p_{M'}(k) = (1-p)^{k}\, p \qquad \text{for } k = 0, 1, 2, \ldots.,$$

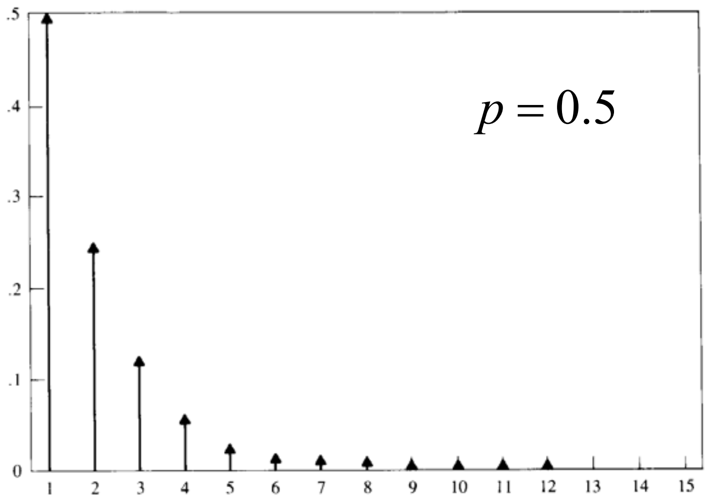
$$E[M'] = \frac{1}{p} \text{-} 1 = \frac{1-p}{p} \qquad \mathrm{VAR}[M'] = \frac{1-p}{p^2}$$

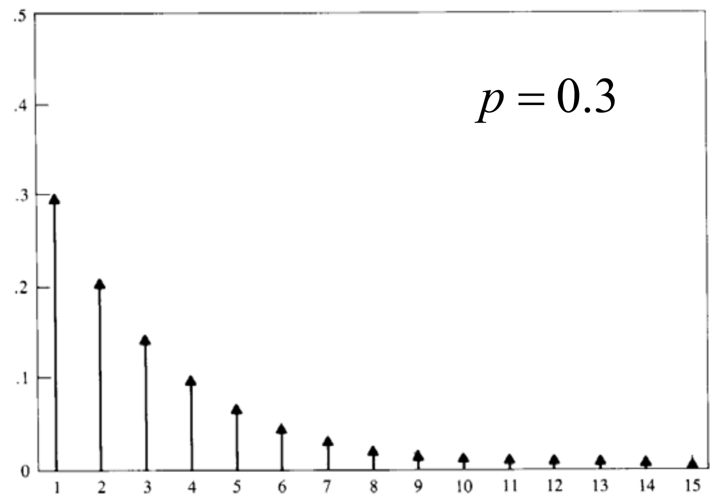
Applications of the Geometric RV

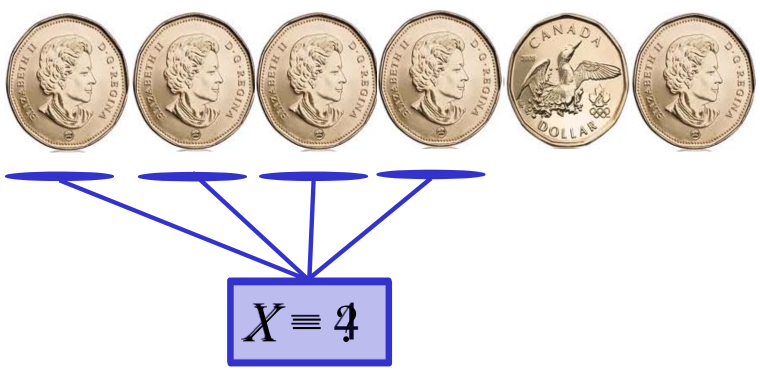
- Number of white dots between successive black dots in a scan of a document
- Number of transmissions required until an error free transmission
- …



Geometric RV: Example PMFs







Memoryless Property of the Geometric RV

- Let M be the number of trials until the first success in a sequence of Bernoulli trials. M is a geometric random variable.

$$P[M = k] = (1-p)^{k-1}\, p \qquad \text{for } k = 1,2,\ldots,$$

- Q: What is the conditional probability that it will take k ***more trials*** until the first success, **given that we have already performed m trials** with no success?
- Solution:

$$P\left[M = k+m \mid M > m\right] = \frac{P\left[\{M = k+m\} \cap \{M > m\}\right]}{P\left[M > m\right]}$$

$$= \frac{(1-p)^{k+m-1}\, p}{(1-p)^{m}} = (1-p)^{k-1}\, p$$

$$= P\left[M = k\right]$$

Memoryless comes from?

Discrete Uniform Random Variable

- A discrete uniform random variable assumes values in a set of consecutive integers with equal probability. $S_X = \{0,1,2,\ldots,M-1\}$

$$\boxed{p_X(k) = \frac{1}{M} \quad \text{for } k \in \{0,1,2,\ldots,M-1\}}$$

- Mean and variance:

$$\boxed{E[X] = \frac{M-1}{2}} \qquad \boxed{Var[X] = \frac{M^2-1}{12}}$$

- Applications
  - Toss of a fair coin or die
  - Spinning of an arrow on a wheel divided into equal intervals (roulette)
  - Mark 6 numbers

Elec 2600H: Lecture 6

- Expectation of a random variable
- Expected value of a function of a random variable
- Variance of a random variable
- Moments of a random variable
- Important discrete random variables
  - Summary of variables you know:
    - Bernoulli
    - Binomial
    - Geometric
    - Discrete Uniform
  - **New random variable: Poisson**
- *Python commands for plotting probability mass functions and generating discrete random variables (FYI)

Poisson Random Variable

- Poisson random variable models the **number of occurrences** of an event in a certain interval of time or space, where events occur **completely at random!**

$$P[N=k]=\frac{\alpha^k}{k!}e^{-\alpha} \qquad \text{for } k=0,1,2,\ldots,$$

The parameter α is the *average number of events in the interval*.

- Mean and variance

$$E[N]=\alpha$$
$$\mathrm{VAR}[N]=\alpha$$

Applications of a Poisson RV

- Number of hits on a website in one hour.
- Number of customer service requests at a service center in one day.
- Number of defects on a semiconductor chip
- Number of photons arriving at a light detector.
- Number of **particles emitted by a radioactive mass** in a fixed time period.
- …

Example

- If X is a Poisson random variable with mean α, show that

$$\boxed{\sum_{k=-\infty}^{\infty} p_X(k) = 1}$$

- **Solution**

  Since the Poisson RV assumes only non-negative values,

$$p_X(k) = 0 \text{ for } k < 0$$

  Thus,

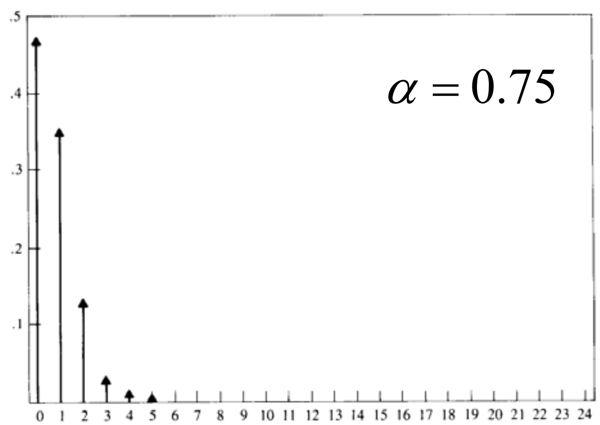
$$\sum_{k=-\infty}^{\infty} p_X(k) = \sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = 1$$

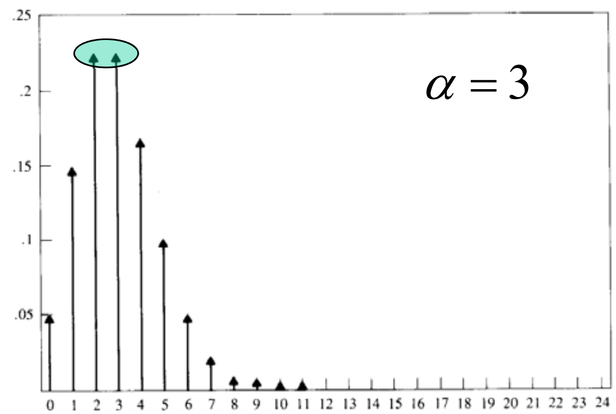
  Since by definition,

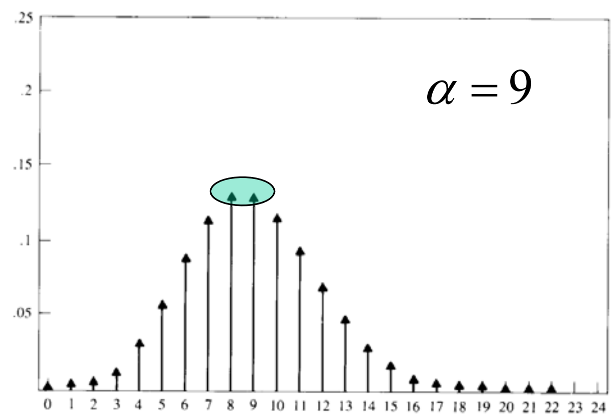
$$e^{\alpha} = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = 1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \ldots$$

Example PMFs

- $P[N=k]$ achieves its maximum value at $\lfloor \alpha \rfloor$, the largest integer smaller than α. If α is a positive integer, then the maximum occurs at α and α -1.







Example

- Requests for telephone connections arrive at a telephone switching office at an **average rate of λ calls per second**. If the number of requests in a time period is **a Poisson random variable**, find the following probabilities:
  - P[no requests in t seconds]
  - P[at least n requests in t seconds]

- Solution

  The number of requests $N(t)$ in t seconds is Poisson with mean $\alpha = \lambda t$. Thus,

$$P[\text{no requests in } t \text{ seconds}] = P[N(t)=0] = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}$$

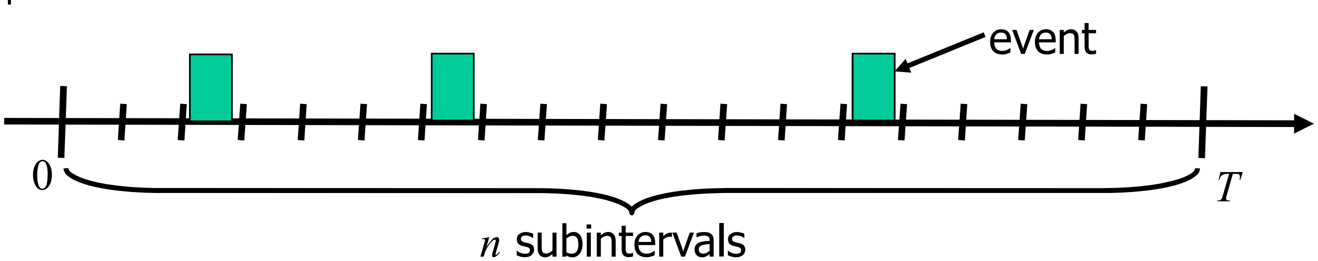
$$P[\text{at least } n \text{ requests in } t \text{ seconds}] = P[N(t) \geq n] = 1 - P[N(t) < n] = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}$$

Poisson Approximation to the Binomial RV

- For **large n and small p**, the pmf of the binomial RV can be approximated by the pmf of the Poisson RV for $\alpha = np$:

$$\binom{n}{k} p^k (1-p)^{n-k} \cong \frac{\alpha^k}{k!} e^{-\alpha}$$

- ***Intuitive argument:***
  - The Poisson RV is the number of occurrences of an event in an interval T, where α is the average number of occurrences.
  - Split the interval into $n > \alpha$ subintervals.
  - Assume that at most one event can occur in each subinterval with probability $p = \alpha/n$ (obtained by the relative frequency interpretation of the probability)
  - As $n \to \infty$,
    - p becomes small
    - The assumption that at most one event can occur in the sub-interval becomes better and better



Mathematical Justification

- For a binomial random variable with $p = \alpha/n$:
  - Consider p_0, the probability of no successes:

$$p_0 = (1-p)^n = \left(1-\frac{\alpha}{n}\right)^n \to e^{-\alpha} \quad \textbf{as } n \to \infty$$

  - Consider the ratio

$$\frac{p_{k+1}}{p_k} = \frac{\binom{n}{k+1} p^{k+1} (1-p)^{n-k-1}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{\frac{n!}{(k+1)!(n-k-1)!} p}{\frac{n!}{k!(n-k)!} (1-p)}$$

$$= \frac{(n-k)p}{(k+1)(1-p)} = \frac{\left(1-\frac{k}{n}\right)\alpha}{(k+1)\left(1-\frac{\alpha}{n}\right)}$$

$$\overset{n\to\infty}{=} \frac{\alpha}{k+1}$$

Mathematical Justification (cont.)

- To summarize: in the limit, $p_0 = e^{-\alpha}$ and $p_{k+1} = \dfrac{\alpha}{k+1} p_k$

- By induction: $p_0 = e^{-\alpha}$

$$p_1 = \frac{\alpha}{1} p_0 = \frac{\alpha}{1} e^{-\alpha}$$

$$p_2 = \frac{\alpha}{2} p_1 = \frac{\alpha^2}{2 \cdot 1} e^{-\alpha}$$

$$p_3 = \frac{\alpha}{3} p_2 = \frac{\alpha^3}{3 \cdot 2 \cdot 1} e^{-\alpha}$$

$$\vdots$$

$$p_k = \frac{\alpha^k}{k!} e^{-\alpha}$$

Poisson pmf!

Example 3.11

The probability of a bit error in a communication line is 10^{-3}. Find the probability that a block of 1000 bits has five or more errors. Bit errors are independent.

Solution

The number of errors, N, is given by a binomial random variable with parameters $n = 1000$ and $p = 10^{-3}$.

Because computing binomial probabilities is very difficult for large n, we use the Poisson approximation with $\boldsymbol{\alpha = np} = 1000 \times 10^{-3} = 1.$

Thus, $P[N \geq 5] = 1 - P[N < 5] = 1 - \sum_{k=0}^{4} \frac{\alpha^k}{k!} e^{-\alpha}$

$$= 1 - e^{-1}\left\{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\right\} = 0.00366$$

Elec 2600H: Lecture 6

- Expectation of a random variable
- Expected value of a function of a random variable
- Variance of a random variable
- Moments of a random variable
- Important discrete random variables
  - Summary of variables you know:
    - Bernoulli
    - Binomial
    - Geometric
    - Discrete Uniform
  - New random variable: Poisson
- ***Python commands for plotting probability mass functions and generating discrete random variables (FYI)**

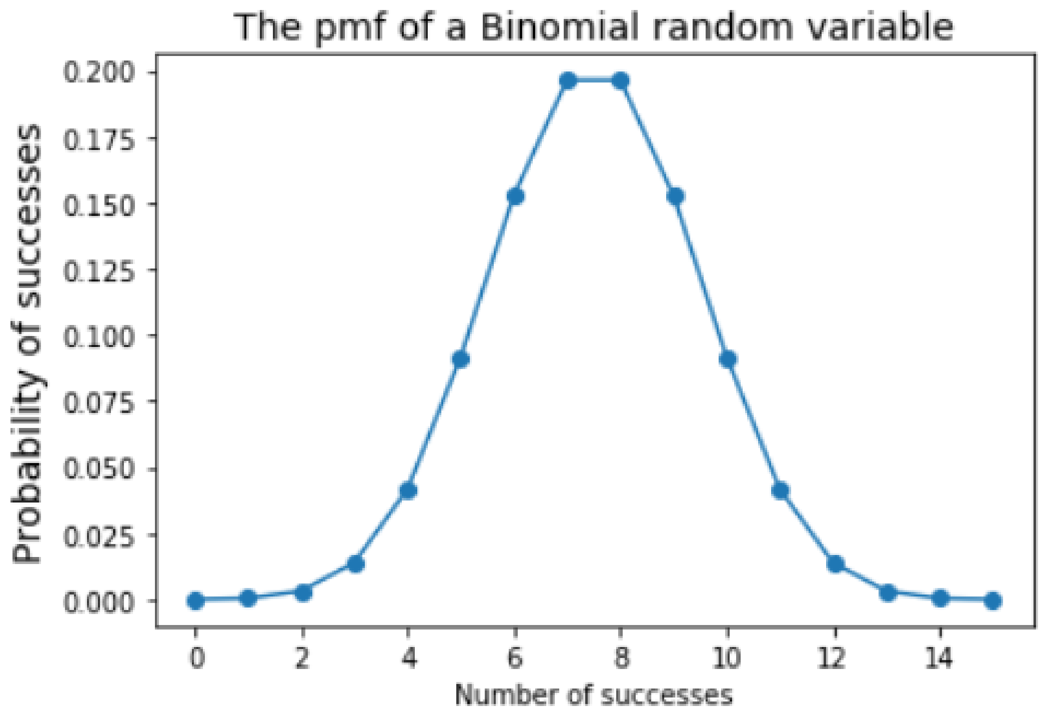
Python Functions for Plotting PMFs

```
import numpy as np #import the numpy package for array object, linear
algebra and random numbers
import scipy.stats #import statistical functions
import matplotlib.pyplot as plt #import the plotting package

n=15 #parameter of a Binomial distribution
p=0.5 #parameter of a Binomial distribution
x=np.arange(0,16) #x values
binomial=scipy.stats.binom.pmf(x,n,p) #values of the binomial pmf
plt.plot(x,binomial,'o-') #generate plot
plt.title('The pmf of a Binomial random variable',fontsize=14)
plt.xlabel('Number of successes')
plt.ylabel('Probability of successes', fontsize=14)

# other pmfs
y = scipy.stats.geom.pmf(x,p) #geometric RV, x starts from 1 not 0
y = scipy.stats.poisson.pmf(x,a)  #Poisson RV
y = scipy.stats.randint.pmf(low,high)#discrete uniform on [low,high-1]
```

Python Functions for Plotting PMFs



Python: Generating Samples

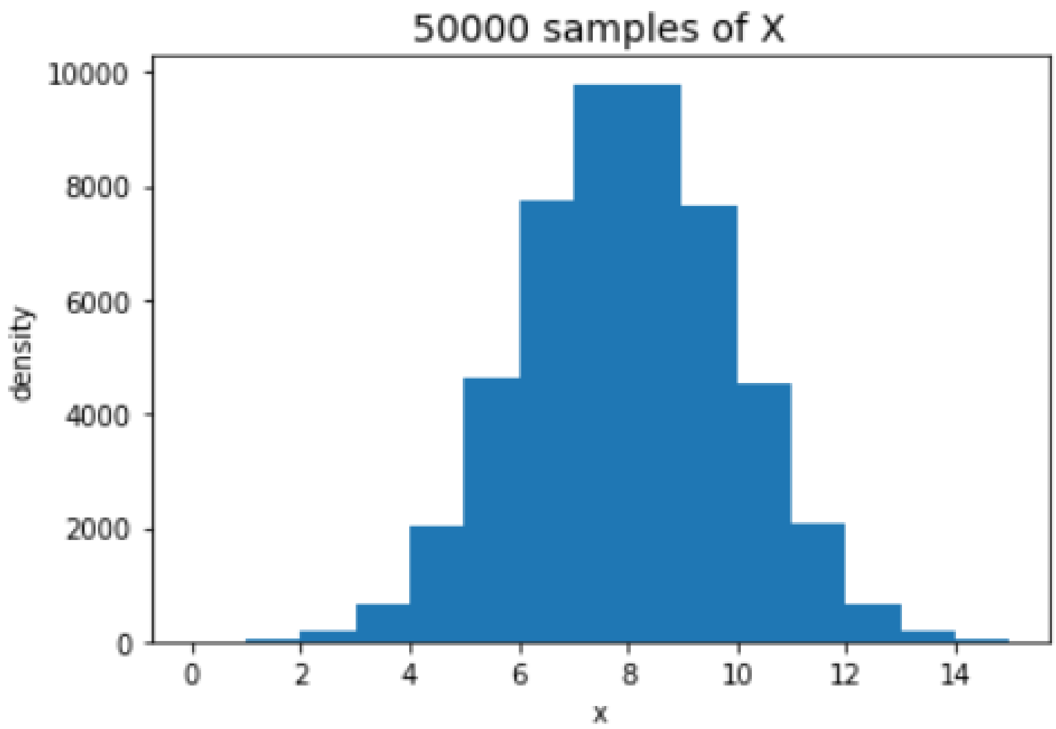
```
import numpy as np #import the numpy package for array object, linear algebra and random numbers
import scipy.stats #import statistical functions
import matplotlib.pyplot as plt #import the plotting package

binom_sim=scipy.stats.binom.rvs(n=15,p=0.5,size=50000)#generate samples follow Binomial distribution with parameters 15 and 0.5
plt.hist(binom_sim,bins=15) #generate plot
plt.title('50000 samples of X',fontsize=14)
plt.xlabel('x')
plt.ylabel('density')



% other pmfs
y = scipy.stats.geom.rvs(p=0.3,size=n)#geometric RV, X starts from 1 not 0
y = scipy.stats.poisson.rvs(mu=2,loc=0,size=n) #poisson RV
y = scipy.stats.randint.rvs(low,high,size=n);  #discrete uniform on [low,high-1]
```

Python Functions for Plotting PMFs



Python Classes for Important Discrete Random Variables

Distribution	Object
Binomial: $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$	**scipy.stats.binom.pmf(x, n, p)**
Geometric: $p_X(x) = (1-p)^{x-1} p$, $x = 1, 2, \ldots$	**scipy.stats.geom.pmf(x, p)**
Uniform: $p_X(x) = \frac{1}{M}$, $x = 0, 1, \ldots, M-1$	**scipy.stats.randint.pmf(low, high)**
Poisson: $p_X(x) = \frac{\alpha^x}{x!} e^{-\alpha}$, $x = 0, 1, \ldots$	**scipy.stats.poisson.pmf(x, α)**

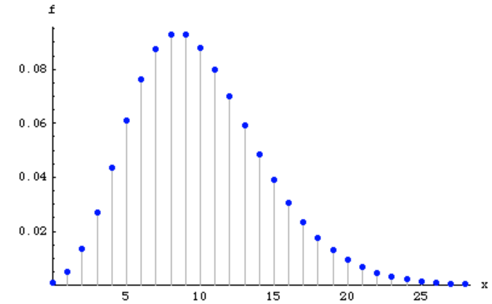
Scipy.stats documentation: https://docs.scipy.org/doc/scipy/reference/stats.html

Important Methods for Discrete Random Variable Classes

Method	Description	Example[1]
rvs	returns **m** samples drawn from the distribution	**y = scipy.stats.binom.rvx(n, p, size = m)**
pmf	returns the pmf at x	**y = scipy.stats.binom.pmf(x, n, p)**

[1] For other distributions, replace scipy.stats.binom with the distribution of interest and set parameters appropriately.

ELEC 2600H: Probability and Random Processes in Engineering



Part II: Single Random Variables

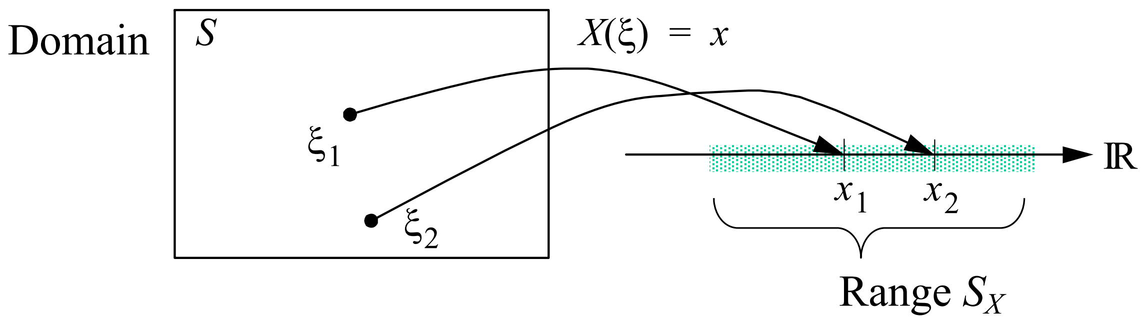
- Lecture 5: Discrete Random Variables
- Lecture 6: Expected Value and Moments; Important Discrete Random Variables
- **Lecture 7: Continuous Random Variables**
- Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs
- Lecture 9: Function of a Random Variable
- Out-of-Class Reading: Conditional PMF and Expectation

Elec 2600H: Lecture 7

- **Single random variables: discrete, continuous and mixed**
  - **Continuous R.V. and Cumulative Distribution Function (CDF)**
  - Probability Density Function (PDF)
  - Conditional CDF’s and PDF’s

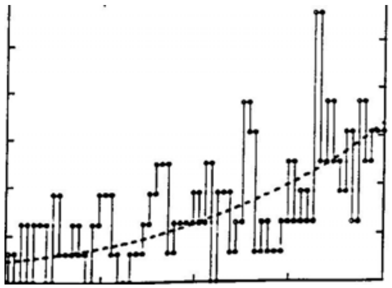
Random Variables: **Review**

- A *random variable* X is a function that assigns a number to every *outcome* of an experiment.
- The function is fixed and deterministic. All randomness in the observed value is due to the underlying experiment



Specifying Probabilities of Random Variables

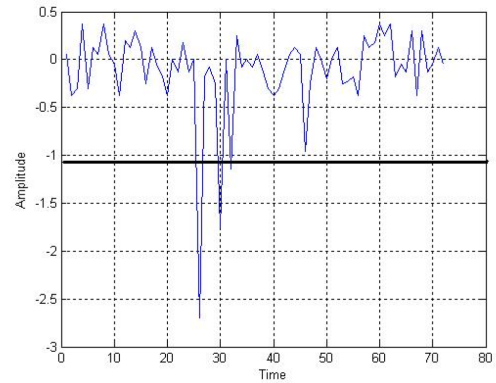
- So far, we have focused on discrete random variables,
  - Possible values come from a countable or finite set.
  - Probability of any event can be computed from the probability mass function.
- But, in many cases, possible values are not naturally restricted to a countable or finite set, but rather come from an **uncountably infinite set**
  - E.g. the unit interval or the real line.
- For random variables like this, we *cannot use the probability mass function*, but we must use one of the following:
  1. **Cumulative distribution function**
  2. Probability density function
  3. Characteristic function

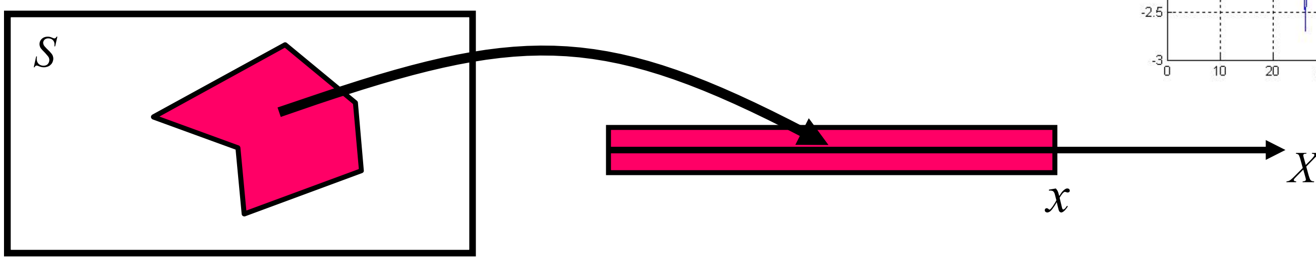


Cumulative Distribution Function (CDF)

- Cdf defined as:

$$F_X(x) = P[X \leq x] \qquad \text{for } -\infty < x < \infty$$





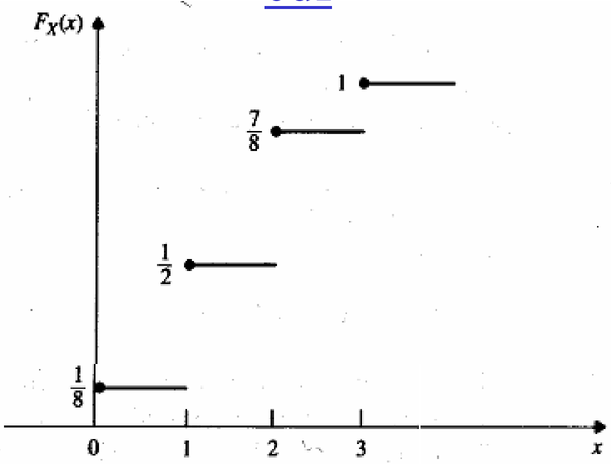
- *Relative frequency interpretation:* if the experiment is performed a large number of times, $F_X(x)$ is the proportion of times that the value of X ended up less than or equal to x.
- Although $F_X(x)$ gives the probabilities of only one type of event (semi-infinite intervals), the probability of *any* event of interest can be computed from it using the axioms and corollaries.

Example 4.1: Three Coin Tosses

- Let X be the number of heads observed in three tosses of a fair coin.
- This is a **discrete random variable** that assumes values in $S_X = \{0,1,2,3\}$
- Properties of this cdf:
  - *Piecewise constant*
  - *Can be expressed as sum of unit steps*
  - *Continuous from the right*

X	0	1	2	3
p_X	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

cdf



- Consider the discontinuity at 1.
  - The limit from the **left** is 1/8.
    - For δ a small positive number,

$$F_X(1-\delta) = P[X \le 1-\delta] = P\{0 \text{ heads}\}$$
$$= 1/8$$

  - The limit from the **right** is 1/2.

$$F_X(1+\delta) = P[X \le 1+\delta] = P\{0 \text{ or } 1 \text{ heads}\}$$
$$= 1/2$$

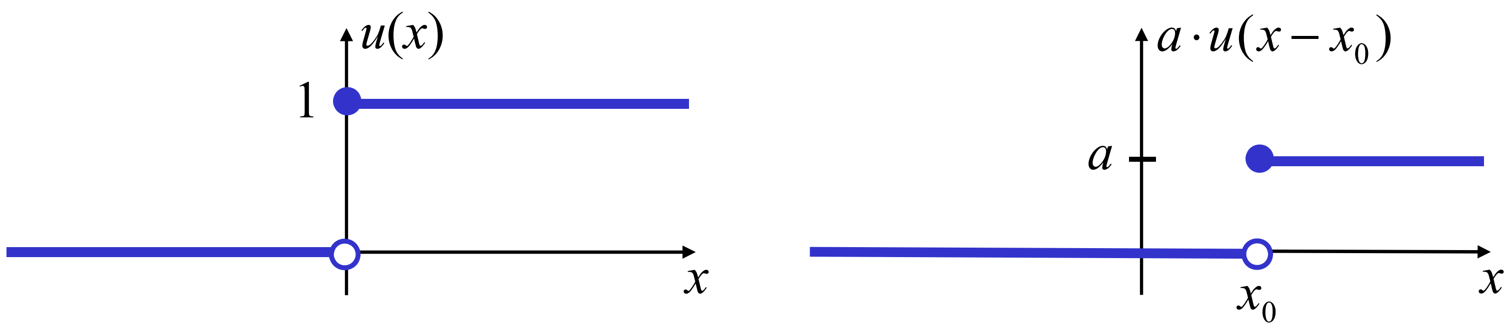
  - CDF value at 1 is the **limit from the right!!**

$$F_X(1) = P[X \le 1] = P\{0 \text{ or } 1 \text{ heads}\}$$
$$= 1/2$$

Unit Step Function

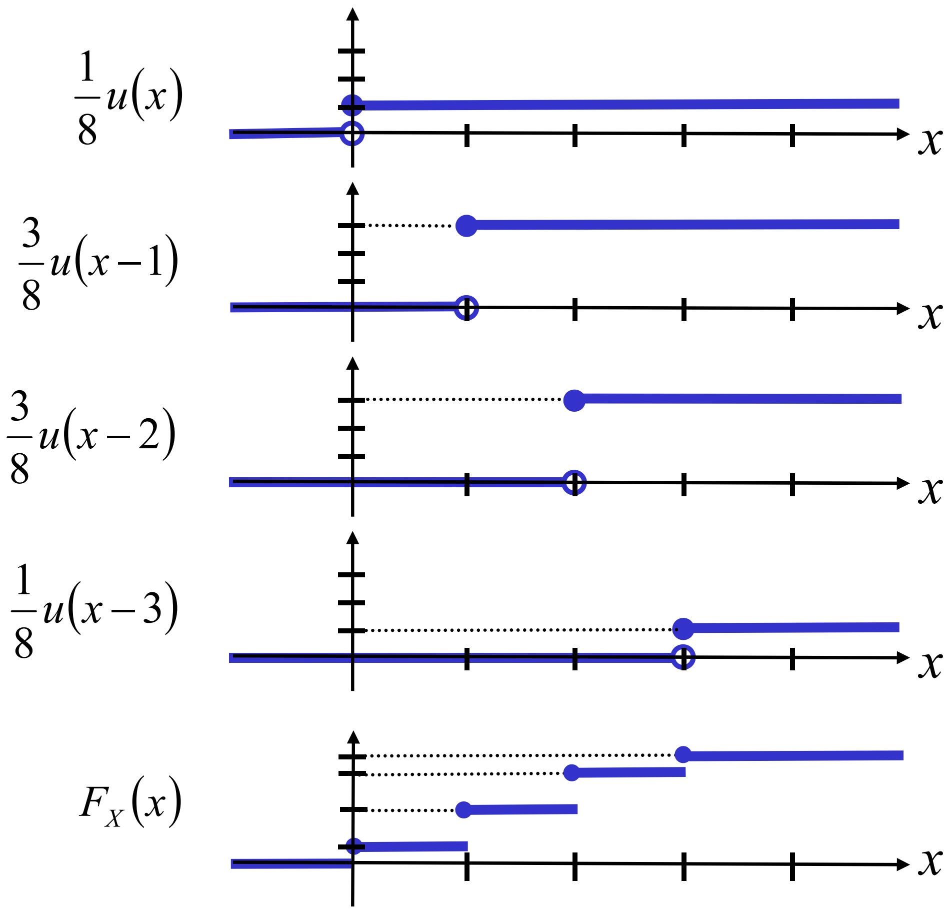
- The unit step function is $u(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$
- Useful to denote functions that are zero for negative values:

  Example: $\left(1 - e^{-\lambda x}\right) \cdot u(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$
- We can build piecewise constant functions as sums of scaled and shifted steps: $a \cdot u(x - x_0)$



CDF = Sum of Step Functions

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3)$$

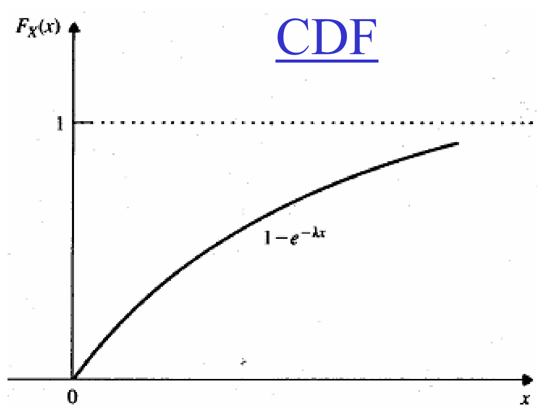


Example: **Continuous Random Variable**

- Waiting time for a taxi at a street corner can be any non-negative real value.
- Commonly used cdf for this situation is ***exponential probability law***:

$$F_X(x) = \left(1 - e^{-\lambda x}\right) \cdot u(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Parameter λ controls the average waiting time (rate function).
- **Larger λ: shorter/longer waiting time?**
- This cdf is continuous everywhere.

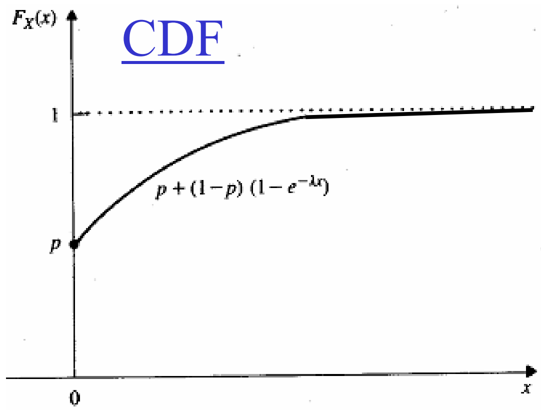


Example: **Mixed Type RV**

- The waiting time, X, at a taxi stand is
  - o Zero if a taxi is already there (with probability p)
  - o Exponentially distributed otherwise.
- Condition on whether taxi is present:

$$P[X \le x | \text{taxi there}] = u(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

$$P[X \le x | \text{no taxi}] = (1 - e^{-\lambda x}) \cdot u(x)$$



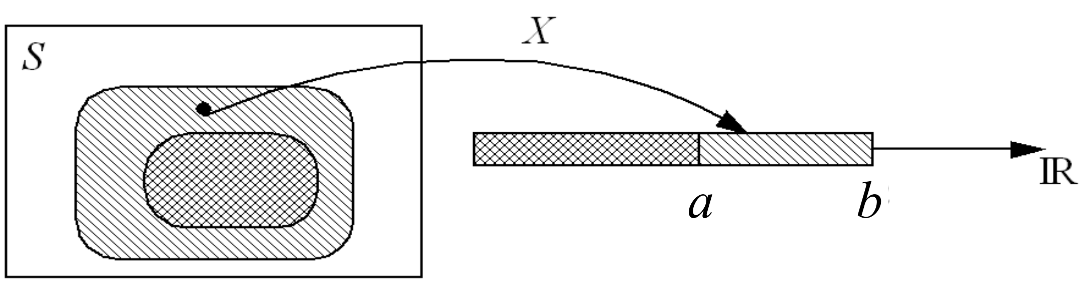
- Using the Total Probability Theorem:

$$F_X(x) = P[X \le x \mid \text{taxi there}]P[\text{taxi there}] + P[X \le x \mid \text{no taxi}]P[\text{no taxi}]$$
$$= u(x) \cdot p + (1 - e^{-\lambda x}) \cdot u(x) \cdot (1-p)$$

- This cdf is ***neither continuous nor piecewise constant***.

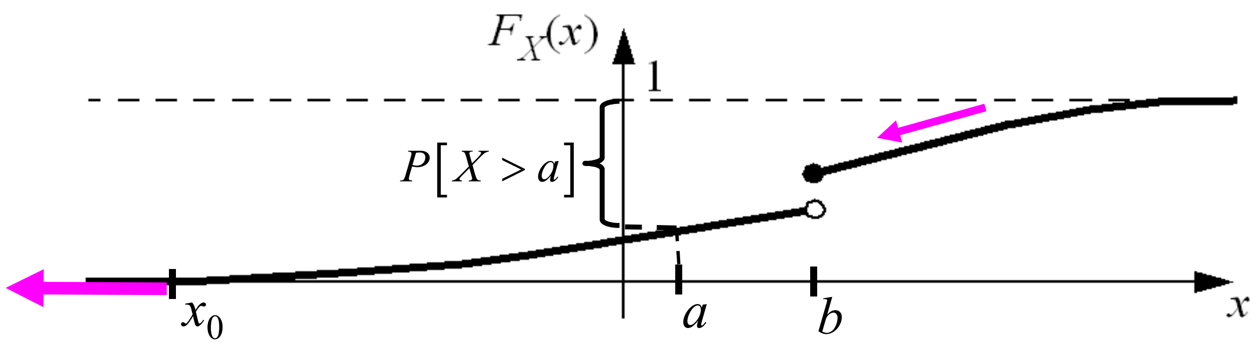
Properties of the CDF (1)

- $\boxed{0 \leq F_X(x) \leq 1}$ (since the cdf is a probability)
- $\boxed{\lim_{x \to \infty} F_X(x) = 1}$ (since every number is $< \infty$)
- $\boxed{\lim_{x \to -\infty} F_X(x) = 0}$ (since no number is $< -\infty$)
- $F_X(x)$ is **non-decreasing**: $\boxed{a < b \quad \Rightarrow \quad F_X(a) \leq F_X(b)}$



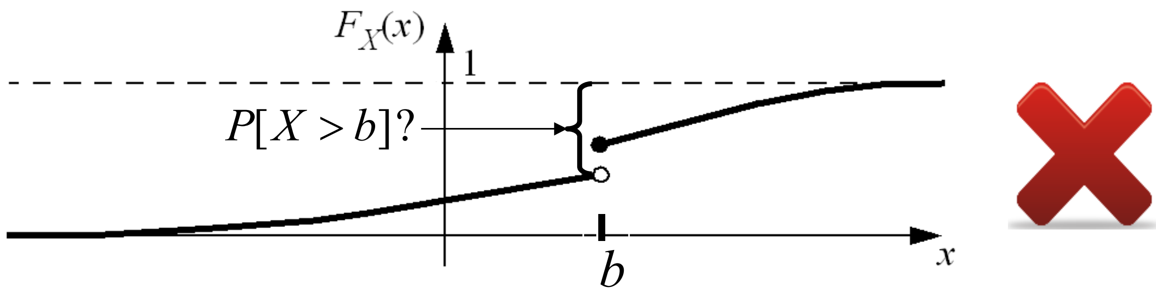
Properties of the CDF (2)

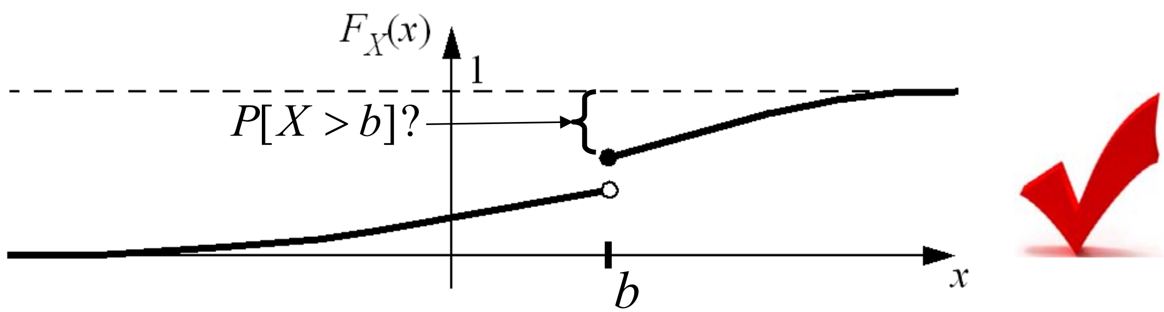
- $F_X(x)$ is **continuous** from the right: $F_X(b) = F_X(b^+) = \lim_{h \to 0} F_X(b+h) \;\text{ for } h > 0$
- If $F_X(x_0) = 0$, then $F_X(x) = 0$ for all $x < x_0$.
- $P[X > a] = 1 - F_X(a)$ is known as the survival function.



Properties of the CDF (2) -- cont'd

Question: We require: $\boxed{P[X > b]}$ Which is correct below?

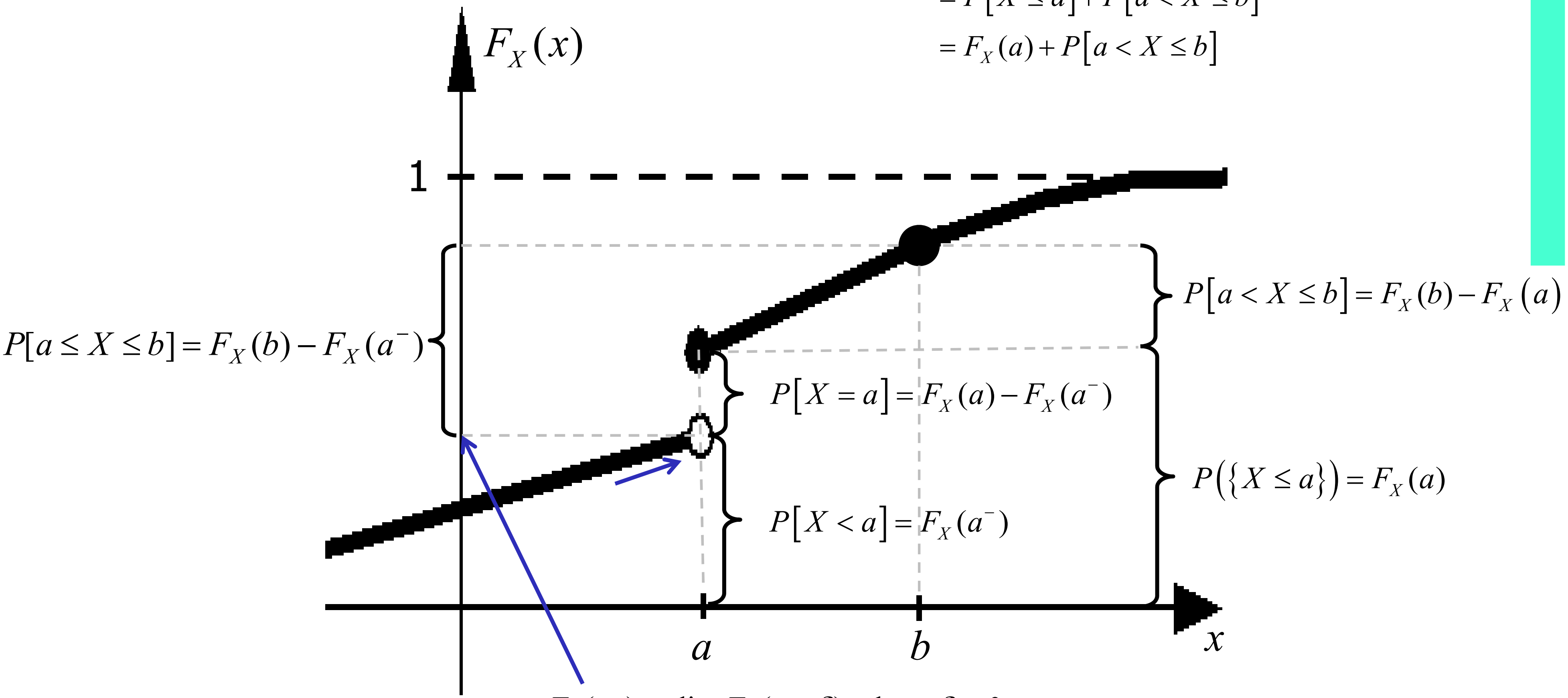




$$P\left[X > b\right] = 1 - P[X \le b] = 1 - F_X\left(b\right)$$

Properties of the CDF (3)

$$F_X(b) = P[X \le b] = P[\{X \le a\} \cup \{a < X \le b\}]$$
$$= P[X \le a] + P[a < X \le b]$$
$$= F_X(a) + P[a < X \le b]$$

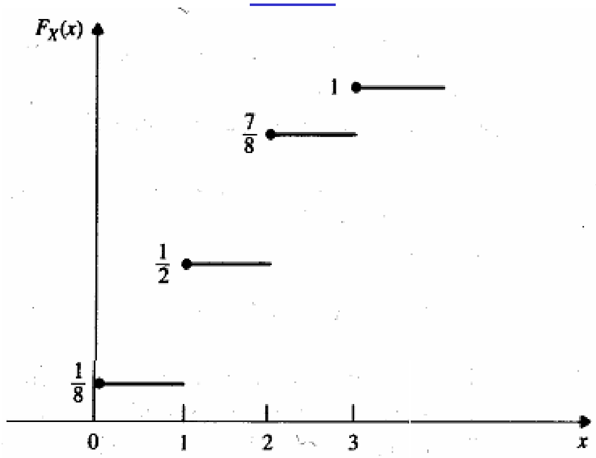


$F_X(a^-) = \lim_{\delta \to 0} F_X(a - \delta)$ where $\delta > 0$

Example 4.4: Three Coin Tosses

- Let X be the number of heads observed in three tosses of a fair coin. Find the probabilities of the following events
  - $A = \{1 < X \leq 2\}$
  - $B = \{0.5 \leq X < 2.5\}$
  - $C = \{1 \leq X < 2\}$

CDF



Solution

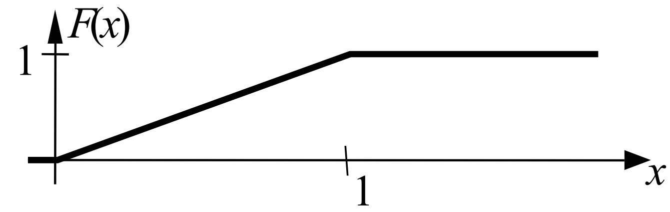
$$P[A] = F_X(2) - F_X(1)$$
$$= 7/8 - 1/2 = 3/8$$

$$P[B] = F_X(2.5^-) - F_X(0.5^-)$$
$$= 7/8 - 1/8 = 6/8$$

$$P[C] = F_X(2^-) - F_X(1^-)$$
$$= 4/8 - 1/8 = 3/8$$

Example 4.5

- Suppose that X has a cdf as shown.



- Find the probabilities of the following events
  - o $A = \{-0.5 < X < 0.25\}$
  - o $B = \{0.3 < X < 0.65\}$
  - o $C = \{|X - 0.4| > 0.2\}$

Solution

$$P[A] = F_X(0.25) - F_X(-0.5)$$
$$= 0.25 - 0 = 0.25$$
$$P[B] = F_X(0.65) - F_X(0.3)$$
$$= 0.65 - 0.3 = 0.35$$
$$P[C] = P[\{X < 0.2\} \cup \{X > 0.6\}]$$
$$= P[\{X < 0.2\}] + P[\{X > 0.6\}]$$
$$= F_X(0.2) + (1 - F_X(0.6)) = 0.2 + 1 - 0.6 = 0.6$$

Three Types of Random Variables: Summary

- A ***discrete random variable*** is a random variable whose cdf is **constant**, except for **jumps at a countable set of points** $x_0, x_1, x_2, \ldots.$
  1. Discrete random variable can only take the values at the jumps.
  2. $p_k = P[X = x_k]$ is the size of the jump at x_k.
  3. The cdf can be written as $F_X(x) = \sum_k p_k u(x - x_k)$

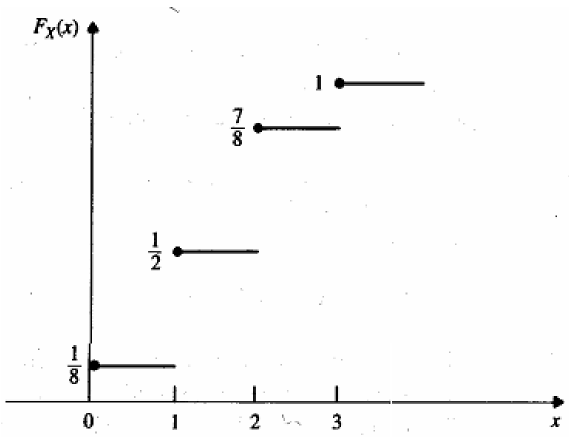
     where $u(x)$ is the unit step function ($u(x) = 1$ if $x \geq 0$ and 0 otherwise).
- A ***continuous random variable*** is a random variable whose cdf is **continuous** everywhere.
  - $P[\{X = x\}] = 0$ everywhere.
- A ***random variable of mixed type*** is a random variable whose cdf has **discrete jumps**, but is **not constant between jumps**.

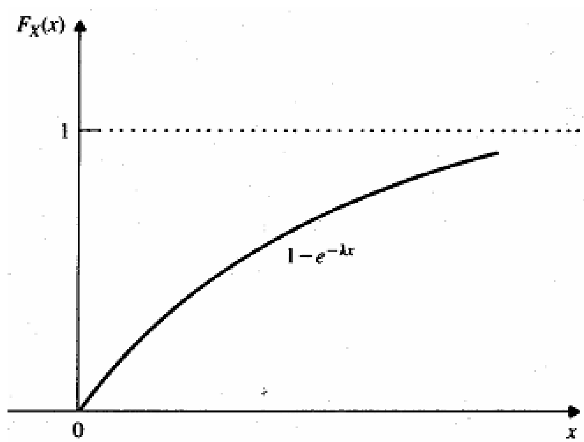
Elec 2600H: Lecture 7

- **Single random variables: discrete, continuous and mixed**
  - Continuous R.V. and Cumulative Distribution Function (CDF)
  - **Probability Density Function (PDF)**
  - Conditional CDF's and PDF's

Intuition

- Values where the cdf is flat are "impossible".
- Values where the cdf is increasing rapidly are "more likely" than values where the cdf is increasing slowly.
- **Thus, the *slope* of the cdf is very informative!!**



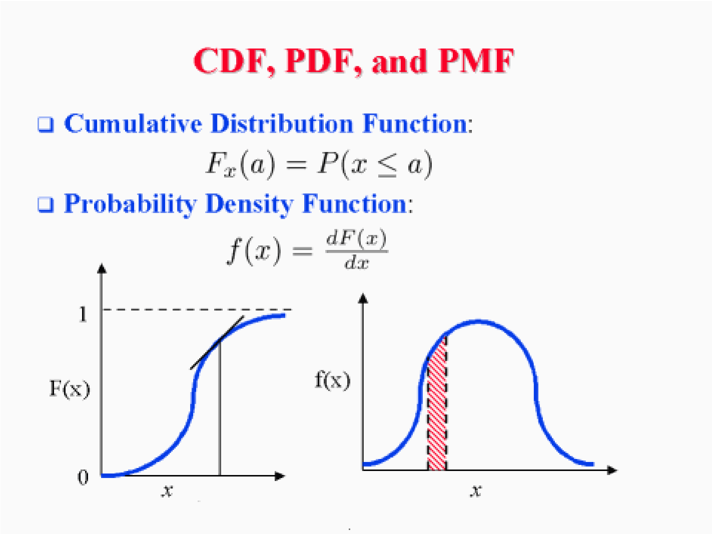


Probability Density Function

- ***Probability Density Function*** **(pdf) is defined as the *derivative of the cdf*:**

$$f_X(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

- Note that the pdf is **NOT** the probability that $X = x$, but is proportional to the probability that X is close to x:

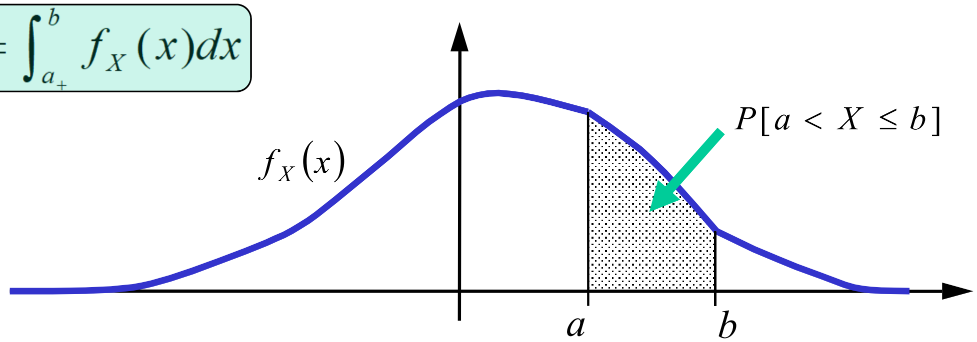


$$\begin{aligned} P[x < X \leq x+h] &= F_X(x+h) - F_X(x) \\ &= \frac{F_X(x+h) - F_X(x)}{h} h \\ &\approx f_X(x) h \quad \text{if } h \text{ is small} \end{aligned}$$

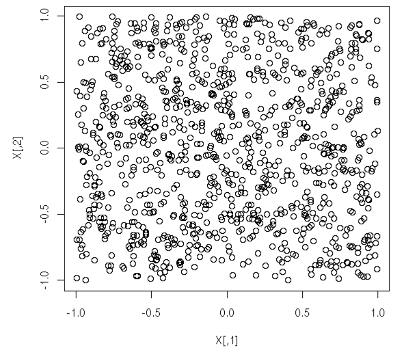
Properties of Probability Density Functions

- $\boxed{f_X(x) \geq 0}$ (since $F_X(x)$ is non-decreasing)
- $\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)dt}$ (the pdf and cdf are equivalent)
- $\boxed{\int_{-\infty}^{+\infty} f_X(t)dt = 1}$ (a good way to check your work!!)
- $P\left[a < X \leq b\right] = F_X(b) - F_X(a_+) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx$

$$\boxed{= \int_{a_+}^{b} f_X(x)dx}$$

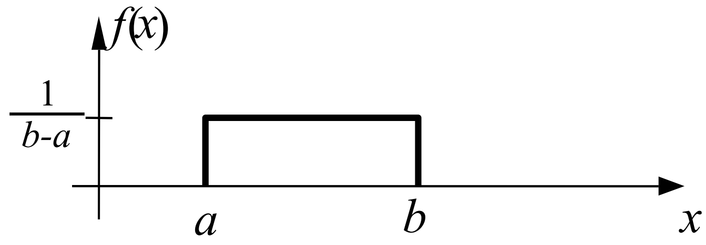


Example 4.6: Uniform Random Variable

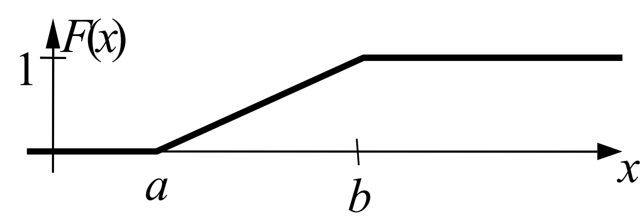


- The uniform r.v. arises in situations where all values in an interval of the real line $[a, b]$ are equally likely to occur.

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



Example 4.7: Exponential RV

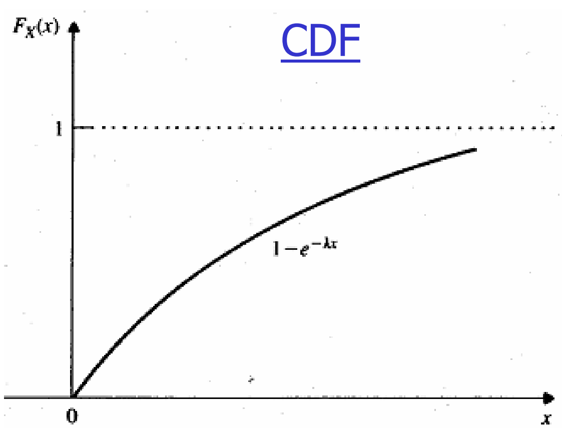
The length of time that passes before the next train arrives can be described with the exponential distribution.

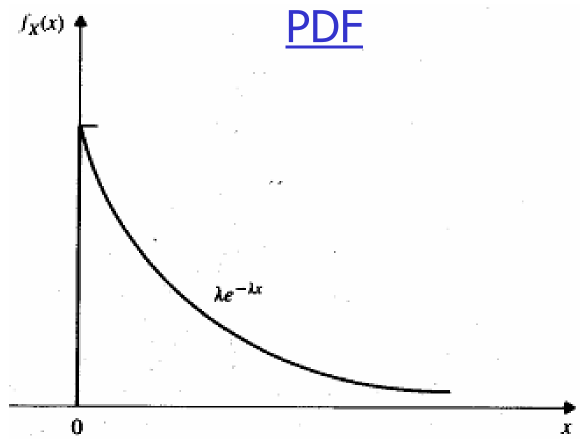
- The exponential probability law has cdf

$$F_X(x) = \left(1 - e^{-\lambda x}\right) \cdot u(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Differentiating, we obtain the following pdf

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



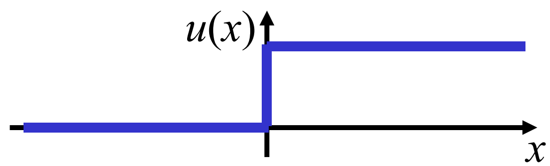


PDF of **Discrete RVs**

- We have seen that the cdf for discrete random r.v.'s contains discontinuous steps.

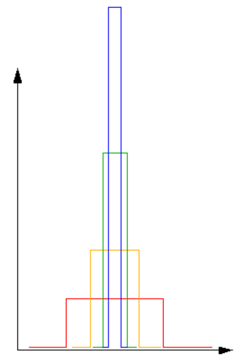
$$F_X(x) = \sum_k p_k u(x - x_k)$$

where $u(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$



- Because the derivative of the unit step is not well defined at the origin, we define its derivative to be a generalized function known as the *delta function*, $d(x)$:
  - $d(x) = 0$ for all $x \neq 0$
  - For any interval (a, b) containing the origin, i.e. $a<0$ and $b>0$, $\int_a^b \delta(x)dx = 1$
  - The exact value of $d(x)$ at $x = 0$ is undefined (infinite)
- This suggests that the **pdf of a discrete r.v.** can be expressed as: $\boxed{f_X(x) = \sum_k p_k \delta(x - x_k)}$

Dirac Delta Function as a Limiting Function



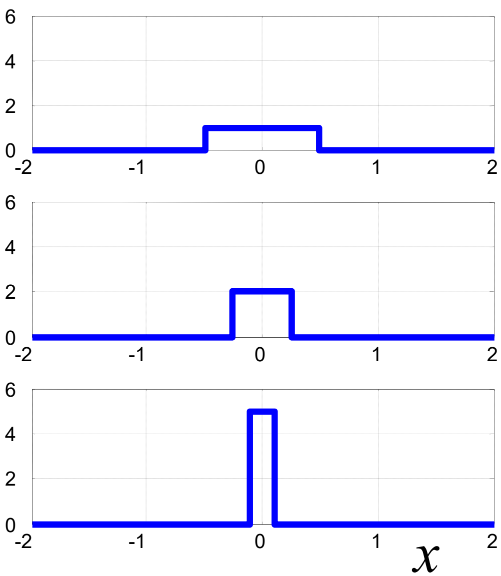
$$\delta(x) = \lim_{A\to\infty} d_A(x) \quad \text{where } d_A(x) = \begin{cases} A & -\frac{1}{2A} \le x \le \frac{1}{2A} \\ 0 & \text{otherwise} \end{cases}$$

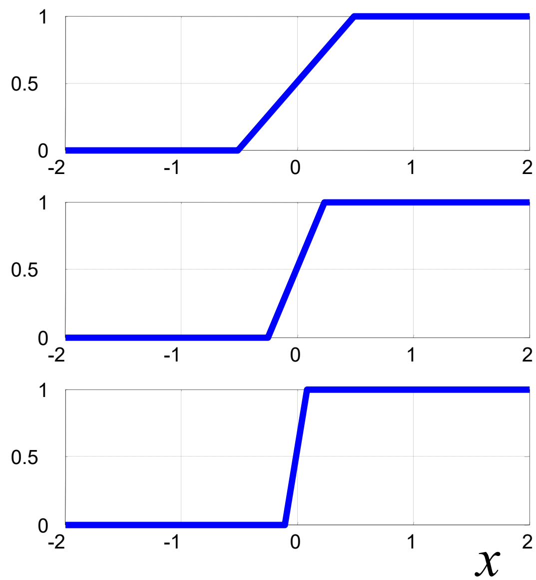
$d_A(x)$ $\qquad$ $\int_{-\infty}^{x} d_A(\eta)\, d\eta$

$A = 1$

$A = 2$

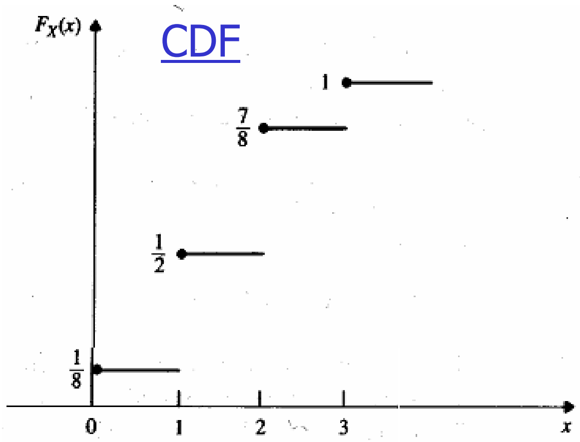
$A = 5$

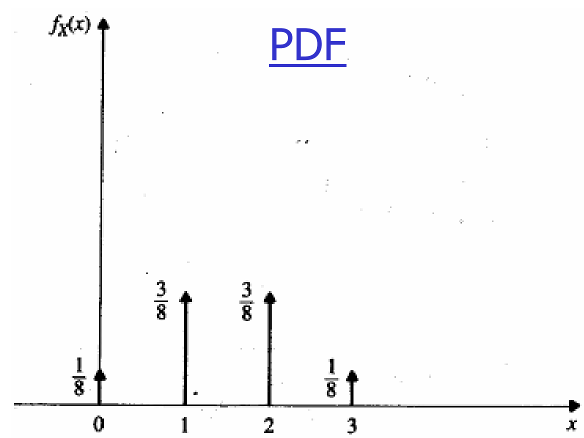




Example 4.9

- Let X be the number of heads observed in three tosses of a fair coin.
- The cdf is $F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x-1) + \frac{3}{8}u(x-2) + \frac{1}{8}u(x-3)$
- Differentiating, the pdf is $f_X(x) = \frac{1}{8}\delta(x) + \frac{3}{8}\delta(x-1) + \frac{3}{8}\delta(x-2) + \frac{1}{8}\delta(x-3)$





Elec2600H: Lecture 7

- **Single random variables: discrete, continuous and mixed**
  - Continuous R.V. and Cumulative Distribution Function (CDF)
  - Probability Density Function (PDF)
  - **Conditional CDF’s and PDF’s**

Conditional CDF's and PDF's

- The conditional cdf a random variable X given an event A is

$$\boxed{F_X(x \mid C) = \frac{P[\{X \le x\} \cap C]}{P[C]}} \quad \text{with } P[C] > 0$$

- The conditional pdf of a random variable X given and event A is

$$\boxed{f_X(x \mid C) = \frac{d}{dx} F_X(x \mid C)}$$

- A conditional cdf or pdf has all the same properties as a regular cdf or pdf.

Example 4.10

The lifetime X of a machine has cdf $F_X(x)$.

Find the conditional cdf given $C = \{ X > t \}$.

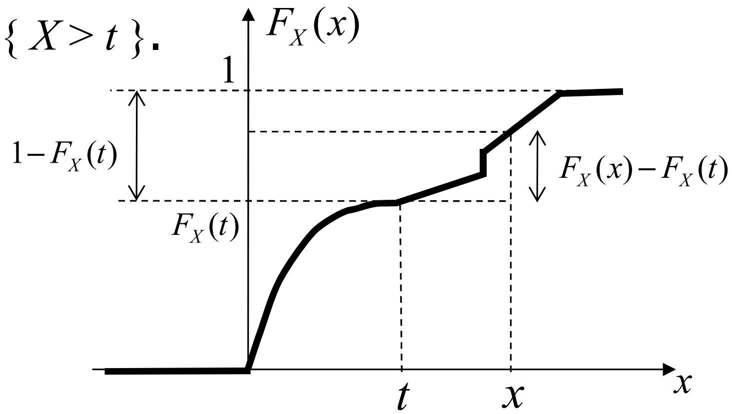
Solution

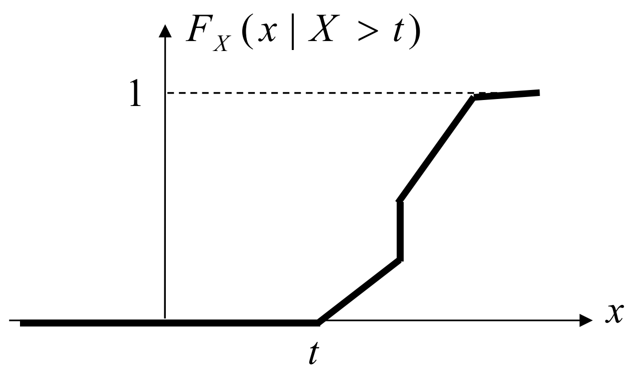
$$F_X(x \mid X > t) = P[X \le x \mid X > t]$$

$$= \frac{P[\{X \le x\} \cap \{X > t\}]}{P[X > t]}$$

$$= \begin{cases} 0 & x \le t \\ \dfrac{P[t < X \le x]}{P[X > t]} & x > t \end{cases}$$

$$= \begin{cases} 0 & x \le t \\ \dfrac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t \end{cases}$$





Example: Device Lifetime

The lifetime X of a machine has an exponential distribution with parameter λ.

$$f_X(x) = \lambda e^{-\lambda x} u(x) \text{ and } F_X(t) = \left(1 - e^{-\lambda t}\right) \cdot u(t)$$

Find the conditional pdf given the event $C = \{ X > t \}$.

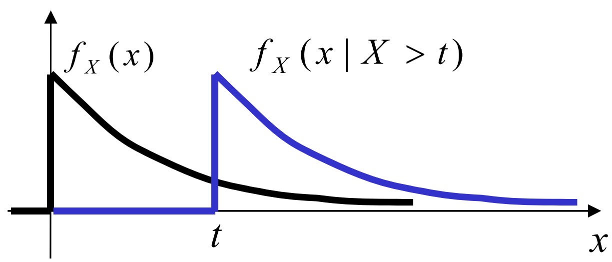
Solution

From previous page, $F_X(x \mid X > t) = \begin{cases} 0 & x \le t \\ \dfrac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t \end{cases}$

Thus,

$$f_X(x \mid X > t) = \begin{cases} 0 & x \le t \\ \dfrac{f_X(x)}{1 - F_X(t)} & x > t \end{cases}$$

$$= \begin{cases} 0 & x \le t \\ \dfrac{\lambda e^{-\lambda x}}{e^{-\lambda t}} = \lambda e^{-\lambda(x-t)} & x > t \end{cases}$$



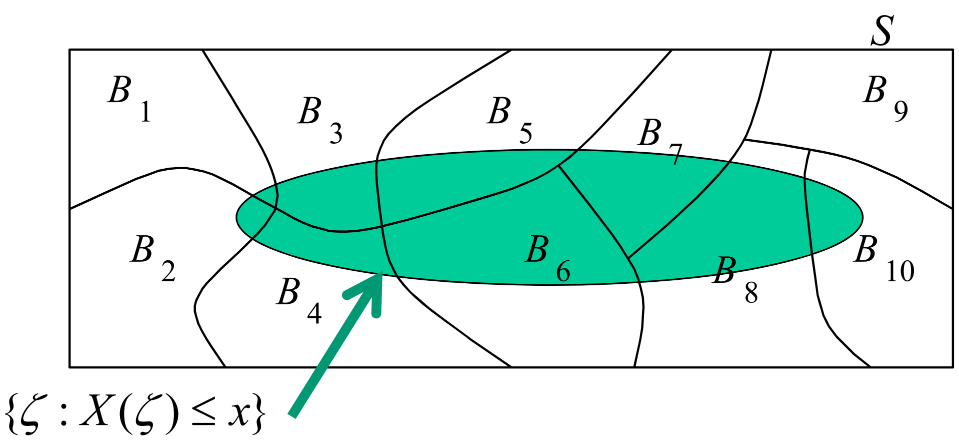
Combining Conditional CDF's and PDF's

- Suppose we have a partition of the sample space $B_1, B_2, \ldots B_n$
- The theorem on total probability allows us to combine conditional cdf's

$$F_X(x) = P[X \le x] = \sum_i P[X \le x | B_i]P[B_i] = \sum_i F_X(x|B_i)P[B_i]$$

- Differentiating with respect to x,

$$f_X(x) = \sum_i f_X(x|B_i)P[B_i]$$



Example: Device Lifetimes

- A production line yields two types of devices
  - Type 1 (**iPod**): short lifetimes exponentially distributed with parameter λ_1
  - Type 2 (**iPad**): long lifetimes exponentially distributed with parameter λ_2
- Suppose the devices occur with probabilities α (Type 1) and (1-α) (Type 2).
- Find the pdf of the lifetime of an arbitrary device, X.

Solution

Let $B_1 = \{\text{device is Type 1}\}$ and $B_2 = \{\text{device is Type 2}\}$.

Since

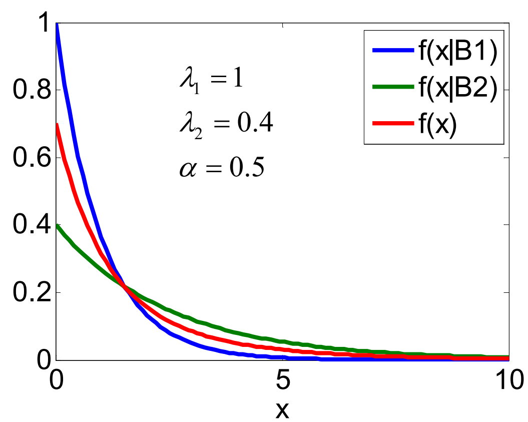
$$f_X(x|B_1) = \lambda_1 e^{-\lambda_1 x} u(x) \qquad P[B_1] = \alpha$$

$$f_X(x|B_2) = \lambda_1 e^{-\lambda_1 x} u(x) \qquad P[B_2] = (1-\alpha)$$

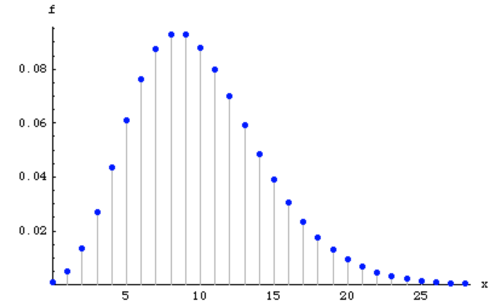
We have that

$$f_X(x) = f_X(x|B_1)P[B_1] + f_X(x|B_2)P[B_2]$$

$$= \left(\alpha\lambda_1 e^{-\lambda_1 x} + (1-\alpha)\lambda_2 e^{-\lambda_2 x}\right)u(x)$$



ELEC 2600H: Probability and Random Processes in Engineering



Part II: Single Random Variables

- Lecture 5: Discrete Random Variables
- Lecture 6: Expected Value and Moments; Important Discrete Random Variables
- Lecture 7: Continuous Random Variables
- **Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs**
- Lecture 9: Function of a Random Variable
- Out-of-Class Reading: Conditional PMF and Expectation

Elec2600H: Lecture 8

- **Expectation of Continuous Random Variables**
- Variance of Continuous Random Variables
- Important Continuous Random Variables

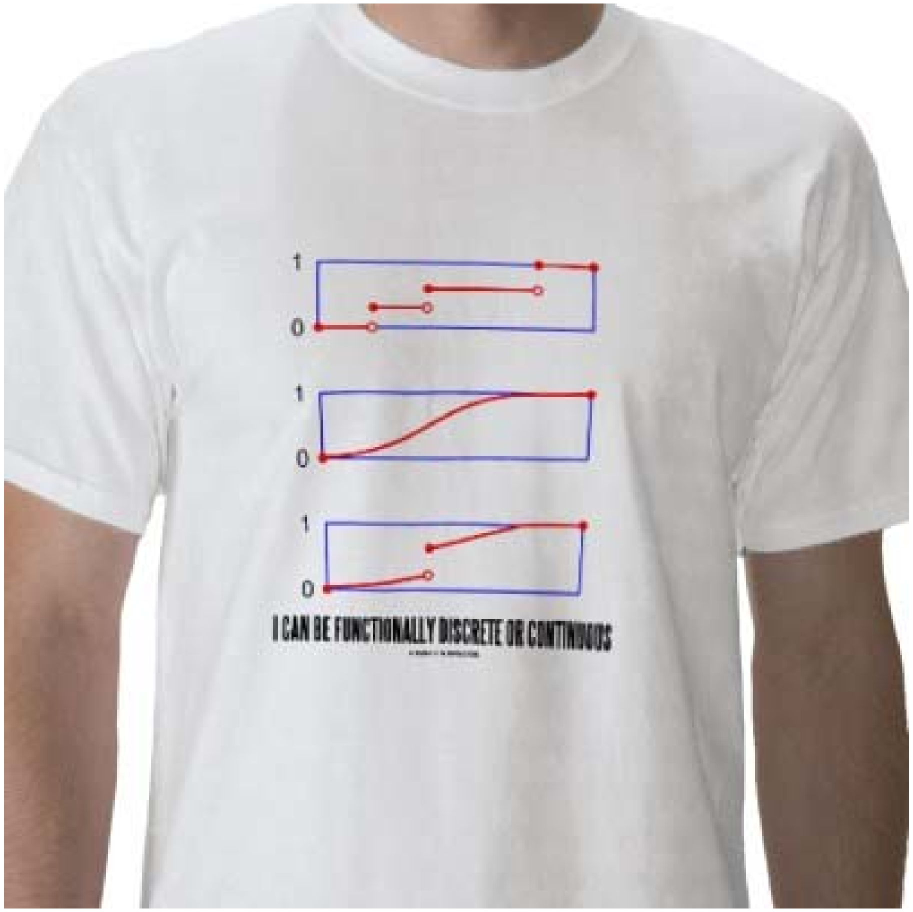
Review: **Expectation**

- Interpretation
  - o The “average” value of a random variable if we repeat the underlying experiment a large number of times.
  - o The limit of the sample mean as the number repetitions of the experiment increases.
  - o Our “best” guess of the value of the random variable. (More on this later.)
- For a discrete RV,

$$E[X] = \sum_k x_k p_X(x_k) \quad \text{where } p_X(x_k) = P[X = x_k]$$

- However, for a continuous RV, $P[X = x] = 0 \text{ for all } x.$

From Discrete to *Continuous* RV



Expectation for Continuous RVs

- The **expected value** or **mean** of a RV is defined as:

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

- The expected value is defined only if the integral converges absolutely.

$$\int_{-\infty}^{+\infty} | x | \, f_X(x)dt < \infty$$

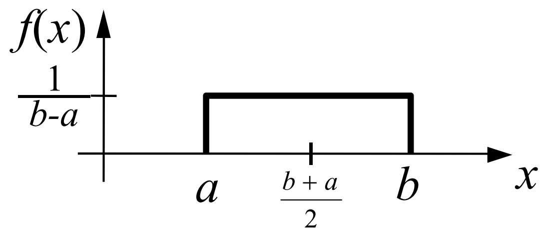
- Expected value often denoted by η, m or μ.
- When X is discrete, the formula reduces to the sum we studied earlier:

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x \sum_k p_X(x_k)\delta(x - x_k)dx \\ &= \sum_k p_X(x_k) \int_{-\infty}^{+\infty} x\delta(x - x_k)dx \\ &= \sum_k x_k p_X(x_k) \end{aligned}$$

Example 4.12: Mean of Uniform Random Variable

- Pdf of a RV X that is uniformly distributed on the interval $[a, b]$ is:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

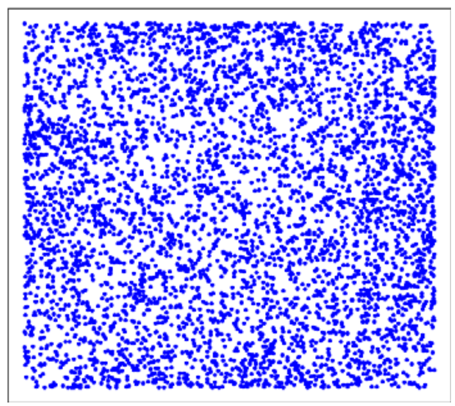


- The expected value is:

$$E[X] = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a} \times \frac{x^2}{2}\bigg|_a^b$$

$$= \frac{1}{b-a} \times \frac{b^2 - a^2}{2} = \frac{(b+a)(b-a)}{2(b-a)}$$

$$= \frac{b+a}{2}$$



Example 4.14: Mean of the Exponential RV

- The pdf of the exponential is $f_X(x) = \lambda e^{-\lambda x} u(x)$ where $u(x)$ is the unit step

- Substituting into the definition and integrating by parts,

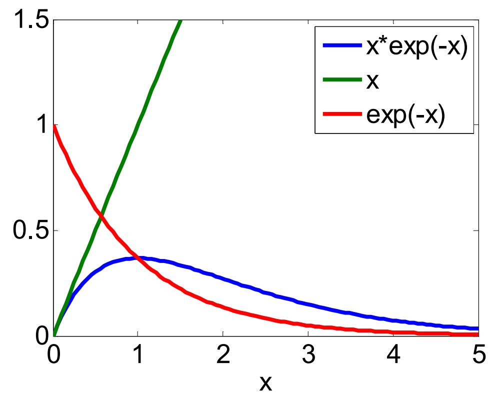
$$
\begin{aligned}
E[X] &= \int_0^\infty x\lambda e^{-\lambda x} dx \\
&= -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx \\
&= -xe^{-\lambda x}\Big|_0^\infty + \frac{e^{-\lambda x}}{-\lambda}\Bigg|_0^\infty \\
&= -0 + 0 + \frac{0}{-\lambda} - \frac{1}{-\lambda} = \frac{1}{\lambda}
\end{aligned}
$$

Integration by parts

$$\int u\,dv = uv - \int v\,du$$

$$u = x \qquad dv = \lambda e^{-\lambda x} dx$$

$$du = dx \qquad v = -e^{-\lambda x}$$



What about the mean of a *FUNCTION* of a continuous random variable…

Expected Value of *a* **Function of a RV**

- Let $Y=g(X)$ where X is a random variable and $g(\cdot)$ is a function. The expected value of Y can be computed from $\boxed{E[Y]=\int_{-\infty}^{\infty} g(x) f_X(x)dx}$
- Although we can compute the expected value from the definition: $E[Y]=\int_{-\infty}^{\infty} y f_Y(y)dy$ using the first formula **avoids the need to find the PDF of Y.**
- This is an extension of the result we proved for **discrete** random variables

$$E[Y]=\sum_k g(x_k)\,p_X(x_k)$$

Example

Suppose that $Y = aX + b$, where X is a random variable and a and b are constants. Find $E[Y]$.

Solution

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= \int (ax+b) f_X(x)\,dx \\ &= a\int x f_X(x)\,dx + b\int f_X(x)\,dx \\ &= \underline{aE[X]+b} \end{aligned}$$

Special cases:

$E[b] = b$ (the mean of a constant is the constant)

$E[aX] = aE[X]$ (scaling the random variable scales its mean)

$E[X+b] = E[X]+b$ (a constant offset shifts the mean by the same amount)

Linearity of the Expectation

- Given a random variable X, constants $a_1, a_2, \ldots a_n$ and functions $g_1, g_2, \ldots g_n$:

$$\boxed{E\left[\sum_i a_i g_i(X)\right] = \sum_i a_i E[g_i(X)]}$$

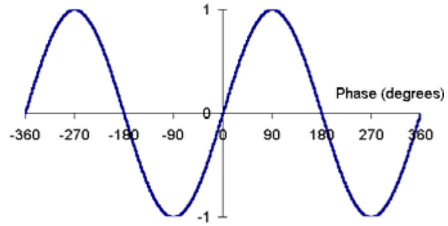
- Intuitively, since expectation, multiplication by a constant and summation are all linear operators, we can switch their order without changing the result.

- Note that it is **generally not true that** $E\left[\sum_i a_i g_i(X)\right] = \sum_i a_i g_i(E[X])$

  unless the g_i are *all linear*.

Example 4.15: Sinusoid with Random Phase

Let $Y = a\cos(\omega t + \Theta)$, where a, ω and t are constants and Θ is uniformly distributed on $[0, 2\pi]$. Find $E[Y]$ and $E[Y^2]$.



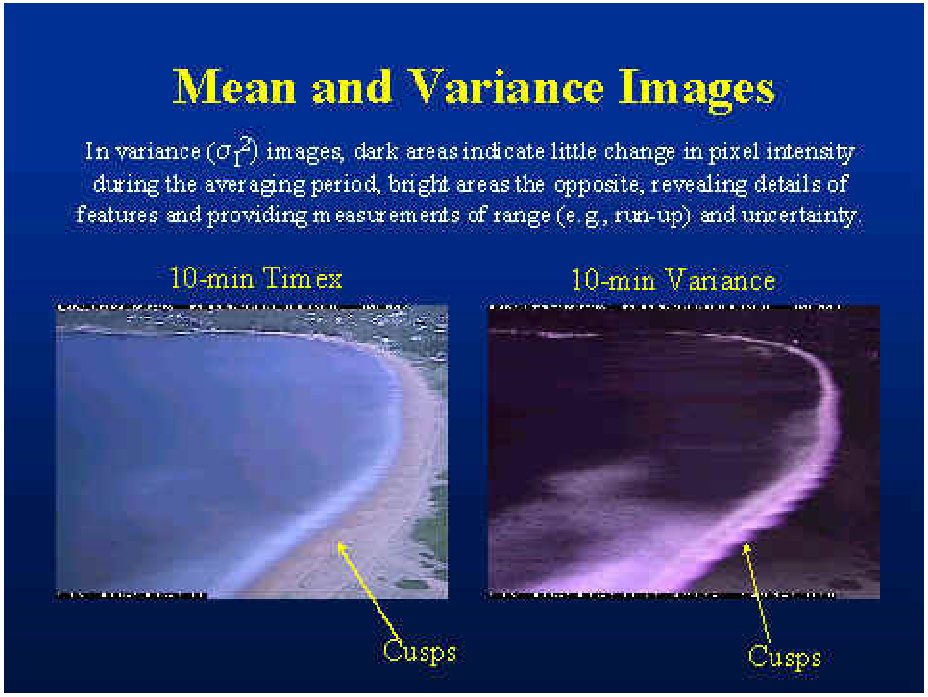
Solution

$$\begin{aligned}
E[Y] &= E[a\cos(\omega t + \Theta)] \\
&= aE[\cos(\omega t + \Theta)] \\
&= a\int_{-\infty}^{\infty} \cos(\omega t + \theta) f_\Theta(\theta)\, d\theta \\
&= a\int_0^{2\pi} \frac{\cos(\omega t + \theta)}{2\pi}\, d\theta \\
&= 0
\end{aligned}$$

$$\begin{aligned}
E[Y^2] &= E[a^2\cos^2(\omega t + \Theta)] \\
&= a^2 E[\cos^2(\omega t + \Theta)] \\
&= a^2 E\left[\frac{1}{2} + \frac{1}{2}\cos(2\omega t + 2\Theta)\right] \\
&= a^2 E\left[\frac{1}{2}\right] + \frac{a^2}{2} E[\cos(2\omega t + 2\Theta)] \\
&= \frac{a^2}{2} + \frac{a^2}{2}\int_0^{2\pi} \frac{\cos(2\omega t + 2\theta)}{2\pi}\, d\theta \\
&= \frac{a^2}{2}
\end{aligned}$$

Elec2600H: Lecture 8

- Expectation of Continuous Random Variables
- **Variance of Continuous Random Variables**
- Important Continuous Random Variables



Variance

- The variance of a random variable measures how much the values of the random variable generally differ from the mean.
- The ***variance*** of a RV, $\mathrm{VAR}[X]$ or σ_X^2, is defined by:

$$\boxed{\mathrm{VAR}[\mathrm{X}] = E[(X - E[X])^2]}$$

- The ***standard deviation*** of a RV, $\mathrm{STD}[X]$ or σ_X, is defined by:

$$\boxed{\mathrm{STD}[\mathrm{X}] = \sqrt{\mathrm{VAR}[\mathrm{X}]}}$$

- All of the properties we studied for the variance of discrete random variables hold true for continuous random variables as well.

Example 4.18: Variance of **Uniform RV**

Let X be uniformly distributed on $[a, b]$. Find $\mathrm{VAR}[X]$.

Solution: $E[X^2] = \int_a^b \frac{x^2}{(b-a)} dx = \frac{1}{b-a}\left(\frac{x^3}{3}\right)\Big|_a^b$

$$= \frac{1}{3}\frac{b^3 - a^3}{b-a} = \frac{1}{3}(b^2 + ab + a^2)$$

Since $E[X]^2 = \frac{(a+b)^2}{4}$,

$$\mathrm{VAR}[X] = E[X^2] - E[X]^2 = \frac{b^2 + ab + a^2}{3} - \frac{a^2 + 2ab + b^2}{4}$$

$$= \frac{(b-a)^2}{12}$$

Example 4.19: Variance of the **Gaussian**

- Since the pdf integrates to 1: $\int_{-\infty}^{\infty} f_X(x)dx = \dfrac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}dx = 1$

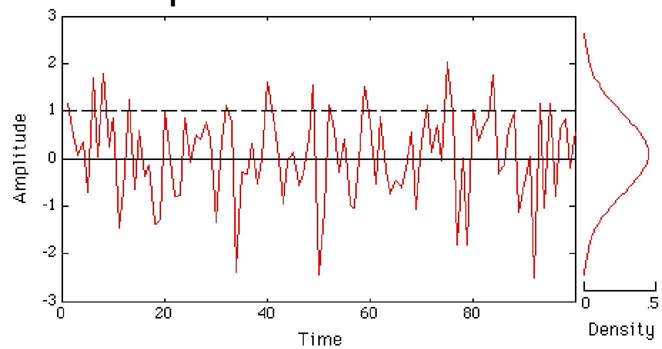
- Multiplying by σ and differentiating with respect to σ:

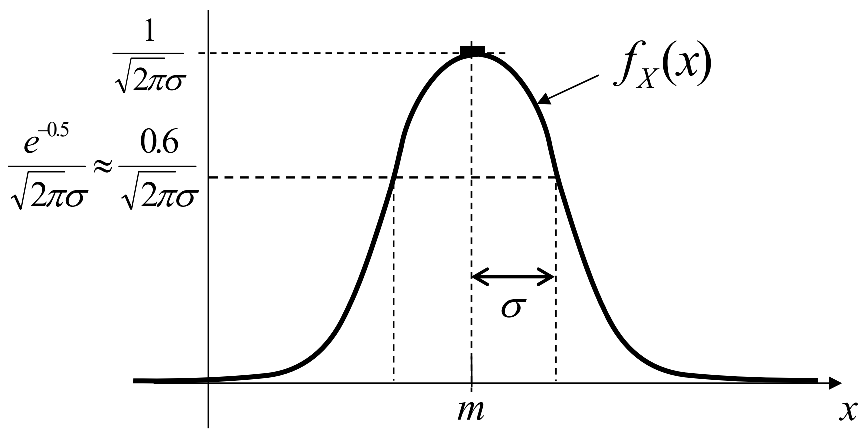
$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}dx = \sigma$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(-\frac{(x-m)^2}{2}\right)\left(\frac{-2}{\sigma^3}\right)e^{-\frac{(x-m)^2}{2\sigma^2}}dx = 1$$

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}}dx = \sigma^2$$

$$\underline{\mathrm{VAR}[X] = \sigma^2}$$





Moments

- The n^{th} (raw) moment of a random variable, X, is defined as

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

- The n^{th} central moment of a random variable, X, is defined as

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

- Important moments:
  - Mean is the 1st moment.
  - Variance is the 2nd central moment.

- The moments are numbers (constants) that summarize different information about the pdf. Under certain conditions, knowing all the moments is equivalent to knowing the pdf.

Minimum Mean Squared Error Estimation

- One way to think of the mean is as a "good" estimate or guess of the value of random variable. But what do we mean by "good"?
- Take any guess, c, of the value of a random variable X. Define the *mean squared error* of the estimate by

$$\boxed{\mathrm{MSE}(c) = E\left[(X-c)^2\right] = \int_{-\infty}^{\infty} (x-c)^2 f_X(x)\,dx}$$

- We obtain the value of c that minimizes the mean squared error by differentiating with respect to c and setting the result equal to zero.

$$\begin{aligned}
\frac{d}{dc}\mathrm{MSE}(c) &= \frac{d}{dc}\int (x-c)^2 f_X(x)\,dx \\
&= -\int 2(x-c) f_X(x)\,dx \\
&= -2\int x f_X(x)\,dx + 2c\int f_X(x)\,dx \\
&= \underline{-2E[X] + 2c = 0}
\end{aligned}$$

Thus, the best guess is $c = E[X]$! Also, the variance is then the *minimum* mean squared error associated with the best guess.

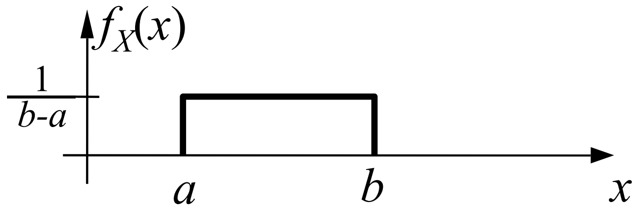
Important Continuous Random Variables

- Expectation of Continuous Random Variables
- Variance of Continuous Random Variables
- **Important Continuous Random Variables**
  - Uniform
  - Exponential
  - Gaussian
  - Laplacian (see text)
  - Gamma (see text)

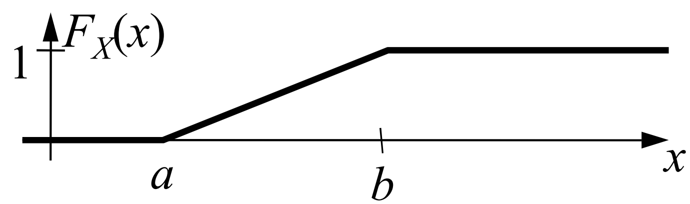
Uniform Random Variable

- Uniform RV arises in situations where all values in an interval of the real line $[a, b]$ are *equally likely* to occur.

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$



$$E[X] = \frac{a+b}{2} \qquad \mathrm{VAR}[X] = \frac{(b-a)^2}{12}$$

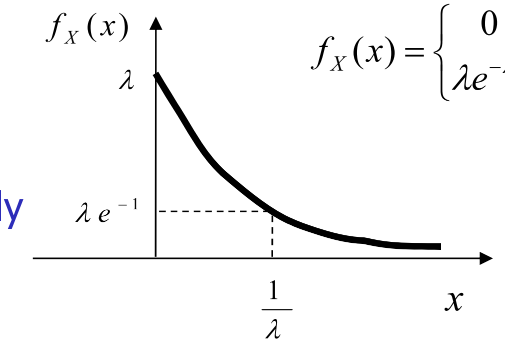
Exponential Random Variable

- The exponential RV can assume values between 0 and infinity.
- Applications
  - o The time between events that occur randomly in time
  - o The time between two typhoons.
  - o The waiting time until the next customer arrives at a bank.
- Mean and variance $E[X]=\frac{1}{\lambda}$

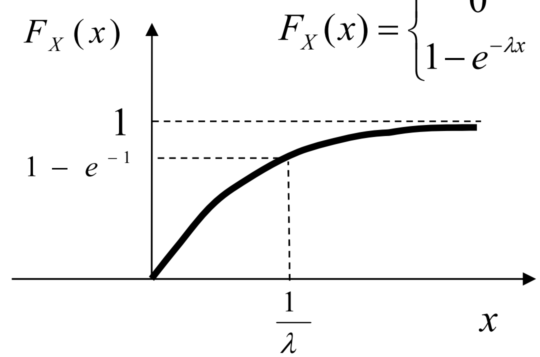
$$\mathrm{VAR}[X]=\frac{1}{\lambda^2}$$

- The parameter λ can be interpreted as the **rate at which events occur**.

$$f_X(x)=\begin{cases}0 & x<0\\ \lambda e^{-\lambda x} & x\geq 0\end{cases}$$



$$F_X(x)=\begin{cases}0 & x<0\\ 1-e^{-\lambda x} & x\geq 0\end{cases}$$



Memoryless Property of the Exponential RV

- Assume that the lifetime of a device is exponentially distributed. The probability that it will continue to work for more than h seconds, given that it has already been in operation for t seconds is

$$P[X > t + h \mid X > t] = \frac{P[\{X > t + h\} \cap \{X > t\}]}{P[X > t]}$$

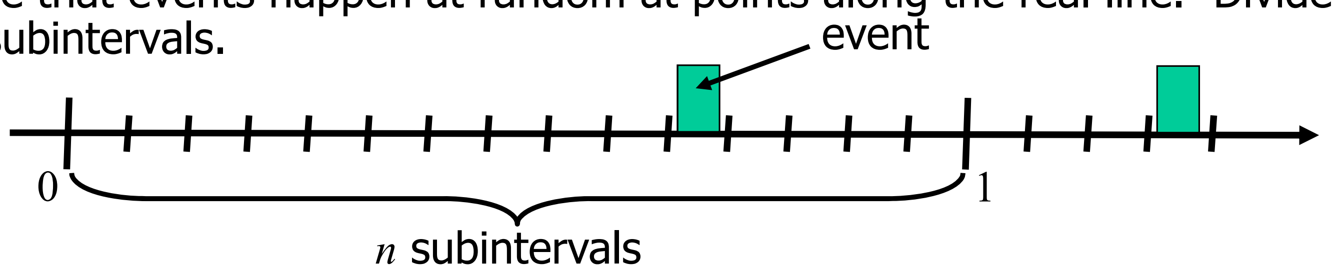
$$= \frac{P[X > t + h]}{P[X > t]} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h}$$

$$= P[X > h]$$

- The conditional probability is independent of how long the part has already been in operation!
- This is the *memoryless property*, since the device has no memory of how long it has been working. In other words, **an old device that is still working will last as long as a new one!**

Geometric and Exponential Random Variables

- Suppose that events happen at random at points along the real line. Divide each unit interval into n subintervals.



- Assume that
  - at most one event can occur per subinterval with probability λ/n
  - events occur independently in different intervals
- The number of subintervals until the first event, M, is a geometric random variable. The actual time is $X = M/n$.

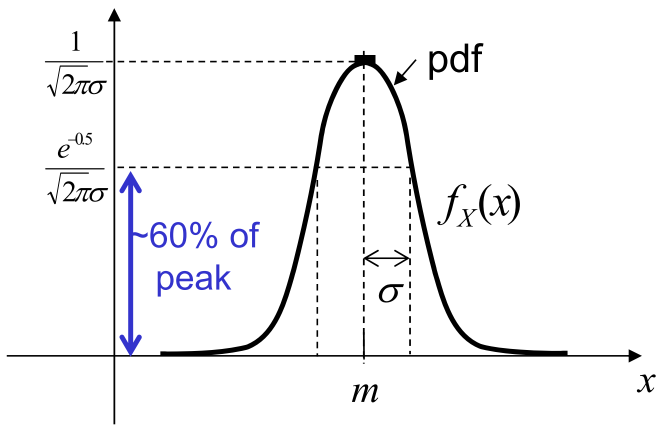
$$P[X > t] = P[M > nt] = (1 - \frac{\lambda}{n})^{nt} \xrightarrow{n \to \infty} e^{-\lambda t}$$

- ***Thus, the exponential random variable is the limiting form of the geometric random variable!***

Gaussian Random Variable

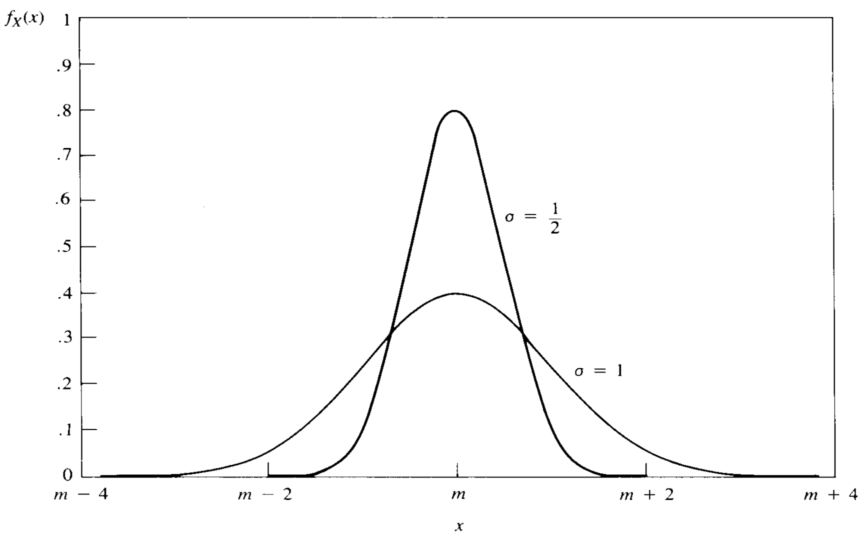
- Gaussian random variable models variables that tend to occur around a certain value, m, the mean.
- This random variable is so common that it is also called the "**normal**" random variable.
- It can assume any value on the real line between -∞ and ∞.
- Applications
  - Noise voltages that corrupt transmission signals.
  - Position of a particle undergoing Brownian motion
  - Voltage across a resistor

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-m)^2}{2\sigma^2}}$$



Gaussian PDF

- The shape of the Gaussian pdf is determined by two parameters
  - the mean: m
  - the standard deviation: σ
- The mean controls the center of the density. The density is symmetric about the mean.
- The standard deviation controls the spread of the density.
- The square of the standard deviation, σ^2, is the variance.

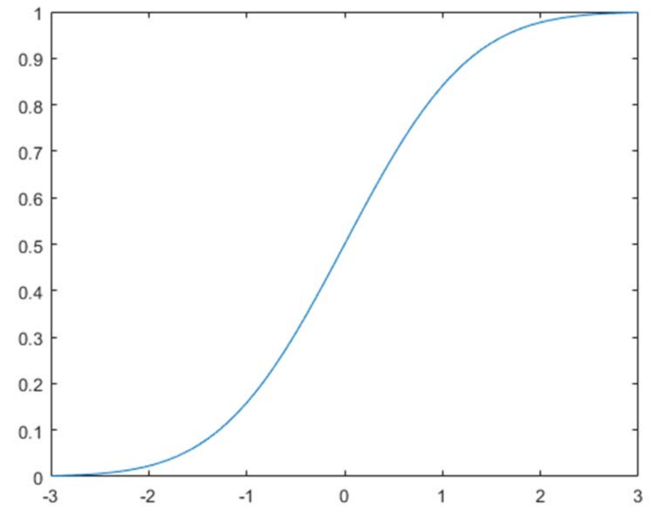


$\Phi(x)$ – CDF of the Normalized Gaussian

- We define the cdf of the "normalized" Gaussian by

$$\Phi(x) = P[X \le x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

  where X has mean $m = 0$ and standard deviation $\sigma = 1$
- **This has no closed form solution.**



- The cdf of a Gaussian RV, Y, with mean m and standard deviation σ, is given by $F_Y(y) = \Phi\left(\frac{y-m}{\sigma}\right)$
- Justification:

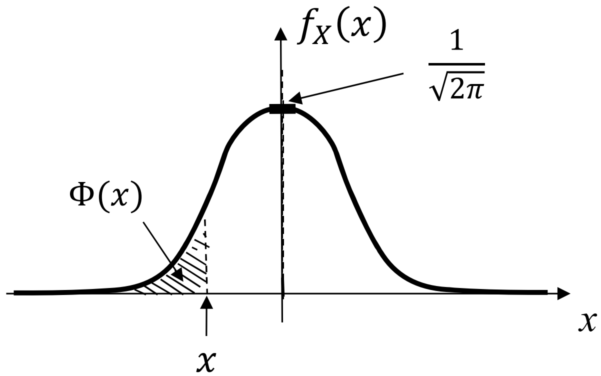
$$F_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{y} e^{-\frac{(\tau-m)^2}{2\sigma^2}}\, d\tau$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{y-m}{\sigma}} e^{-\frac{t^2}{2}} dt$$

$$= \Phi\left(\frac{y-m}{\sigma}\right)$$

Apply change of variables

$$t = \frac{\tau - m}{\sigma}$$



Q-function

- In electrical engineering, we often use the Q-function, rather than the cdf, to express probabilities involving Gaussian RVs.

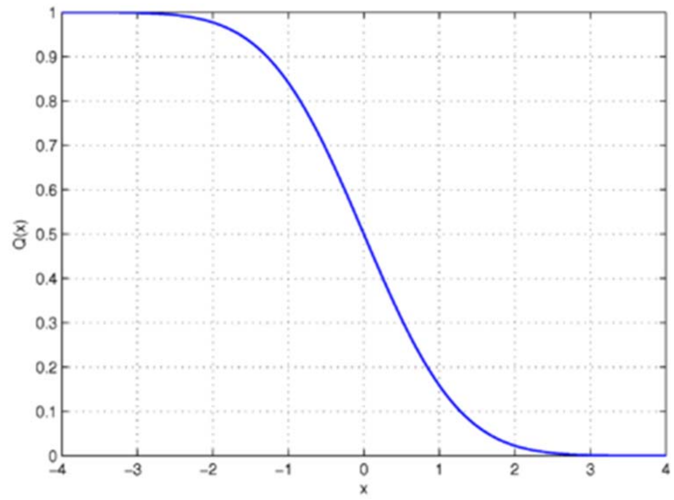
  $$Q(x) = P[X > x]$$

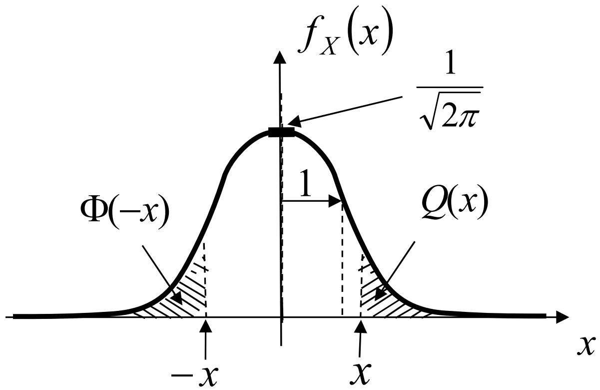
  where X is the normalized Gaussian.

- The Q-function is an example of a *survival function*: the probability that a random variable is greater than a certain value.

- Relationships between Q and Φ:
  - $Q(x) = 1 - \Phi(x)$
  - $Q(x) = \Phi(-x)$

- For a Gaussian RV, Y, with mean m and standard deviation σ, $P[Y > y] = Q\left(\frac{y-m}{\sigma}\right)$





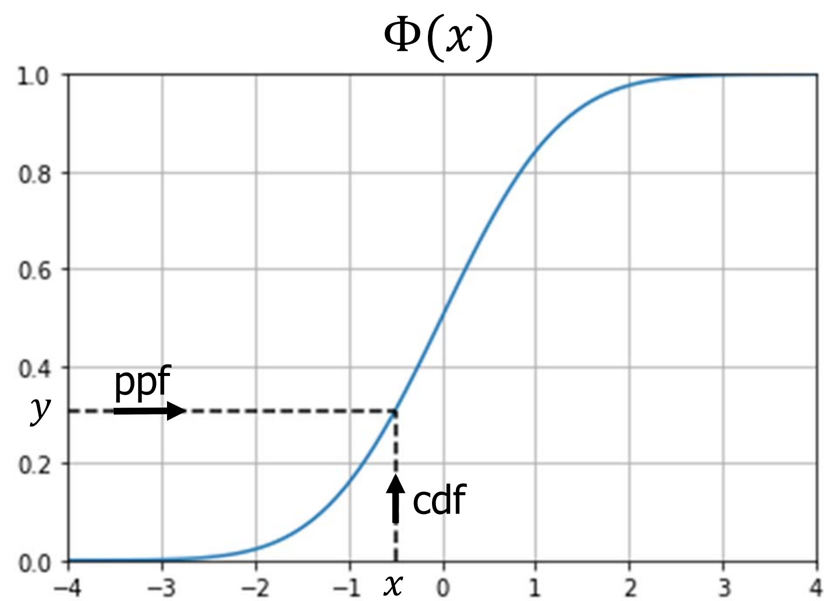
Calculating probabilities involving Gaussians

- The cdf and survival functions of Gaussian random variables have no closed form solution, so they must be computed numerically.
- The **norm** class in the **scipy.stats** package in Python contains methods for computing the cdf, survival and other related functions of Gaussian random variables. The loc parameter sets the mean, and the scale parameter sets the standard deviation

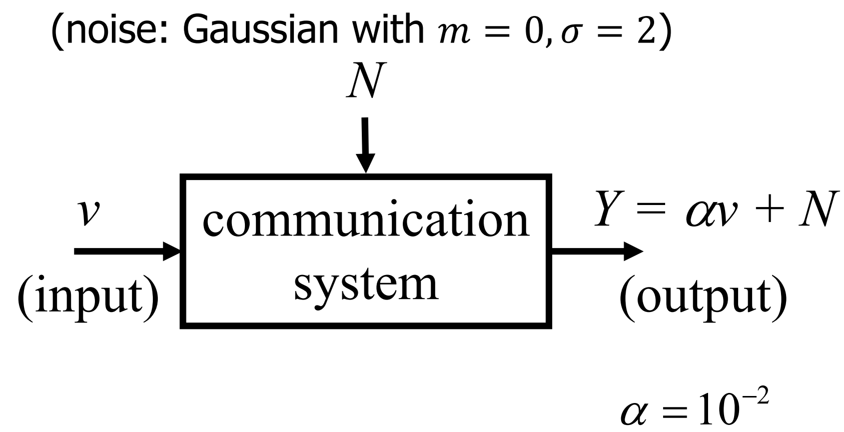
```
from scipy import stats
# two ways to compute the cdf of normalized Gaussian
y = stats.norm.cdf(x)
y = stats.norm.cdf(x, loc=0, scale=1)
# cdf of Gaussian with mean m and standard deviation sd
y = stats.norm.cdf(x, loc=m, scale=sd)

# the point percentage function (inverse of cdf)
x = stats.norm.ppf(y)


y = stats.norm.sf(x)  # Q-function
x = stats.norm.isf(y) # inverse of Q-function
```



Example 4.22



Determine the minimum value of v so that $P[Y < 0] \leq 10^{-6}$

Solution

$$
\begin{aligned}
P[Y < 0] &= P[av + N < 0] \\
&= P[N < -av] \\
&= \Phi\left(\frac{-av-0}{\sigma}\right) \\
&= \Phi\left(-\frac{av}{\sigma}\right)
\end{aligned}
$$

Thus, we must have $\Phi\left(-\frac{av}{\sigma}\right) \leq 10^{-6}$

$$-\frac{\alpha}{\sigma}v \leq \Phi^{-1}(10^{-6})$$

$$v \geq -\frac{\sigma}{\alpha}\Phi^{-1}(10^{-6})$$

Using python, we must have $v \geq$ `- 2 / 1e-2 * stats.norm.ppf(1e-6) = 950.7`

Python Classes for Important Continuous Random Variables

Distribution	**Object**
Exponential: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$	**scipy.stats.expon(scale = $\frac{1}{\lambda}$)**
Gaussian: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$	**scipy.stats.norm(loc = m, scale = σ)**
Uniform: $f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$	**scipy.stats.uniform(loc = a, scale = b-a)**
Laplacian: $f_X(x) = \frac{\alpha}{2} e^{-\alpha\|x\|}$	**scipy.stats.laplace(scale = $\frac{1}{\alpha}$)**
Gamma: $f_X(x) = \begin{cases} \frac{\lambda(\lambda x)^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$	**scipy.stats.gamma(a = α, scale = $\frac{1}{\lambda}$)**

Scipy.stats documentation:
https://docs.scipy.org/doc/scipy/reference/stats.html

Important Methods for Continuous Random Variable Classes

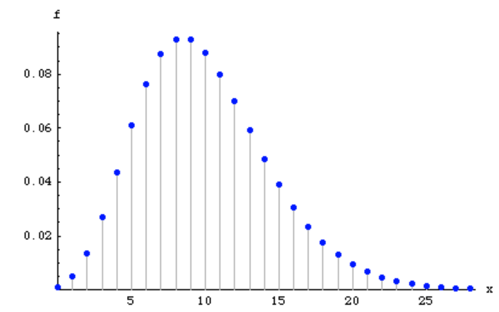
Method	Description	Example[1]
rvs	returns **n** samples drawn from the distribution	y **= scipy.stats.expon.rvs (scale=**$\frac{1}{\lambda}$**, size=**n**)**
pdf	returns the pdf	y **= scipy.stats.expon.pdf (**x**, scale =** $\frac{1}{\lambda}$**)**
cdf	returns the cdf	y **= scipy.stats.expon.cdf (**x**, scale =** $\frac{1}{\lambda}$**)**
ppf	returns the percent point function (the inverse of the cdf)	x **= scipy.stats.expon.ppf (**y**, scale =** $\frac{1}{\lambda}$**)**
sf	returns the survival function, i.e. the probability the random variable is greater than x	y **= scipy.stats.expon.sf (**x**, scale =** $\frac{1}{\lambda}$**)**
isf	returns the inverse of the survival function	x **= scipy.stats.expon.isf (**y**, scale =** $\frac{1}{\lambda}$**)**

One use of the percentage point function is to set plot limits. For example, to ensure that the region covered by your plot includes 98% of the probability:

```
x = np.linspace( scipy.stats.norm.ppf(0.01), scipy.stats.norm.ppf(0.99) )
plt.plot( x, scipy.stats.norm.pdf(x) )
```

[1] For other distributions, replace scipy.stats.expon with the distribution of interest and set parameters appropriately.

ELEC 2600H: Probability and Random Processes in Engineering



Part II: Single Random Variables

- Lecture 5: Discrete Random Variables
- Lecture 6: Expected Value and Moments; Important Discrete Random Variables
- Lecture 7: Continuous Random Variables
- Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs
- **Lecture 9: Function of a Random Variable**
- Out-of-Class Reading: Conditional PMF and Expectation

Elec2600H: Lecture 9

- **Functions (or *Transformations*) of a Random Variable**

Functions/Transforms of a Random Variable

Problem statement:

Given a random variable X with known distribution and a real valued function $g(x)$, such that $Y = g(X)$ is also a random variable. Find the distribution of Y.

Solution methodology:

- The most important thing is to determine *what type* of random variable Y is! Solution approach depends critically upon this.
  - If X is discrete, Y **must** be discrete, irrespective of $g(x)$.
  - If X is continuous, Y may be discrete, continuous or mixed-type. This depends on the form of $g(x)$.
- We will present three different approaches. They are all based on the idea of finding an equivalent events in X for suitably defined events in Y.

Three Approaches

1. If Y is *discrete*, find the pmf at the possible values of Y:

$$p_Y(k) = P[\underbrace{Y = k}_{\text{event in } Y}] = P[\underbrace{\{X : g(X) = k\}}_{\text{equivalent event in X}}]$$

2. If Y is *continuous*, there are two approaches:

   a. Find the cdf of Y:

$$F_Y(y) = P[Y \le y] = P[\{X : g(X) \le y\}]$$

   b. Find the pdf of Y:

$$f_Y(y_0) = \sum_{x:g(x)=y} \frac{f_X(x)}{|g'(x)|}$$

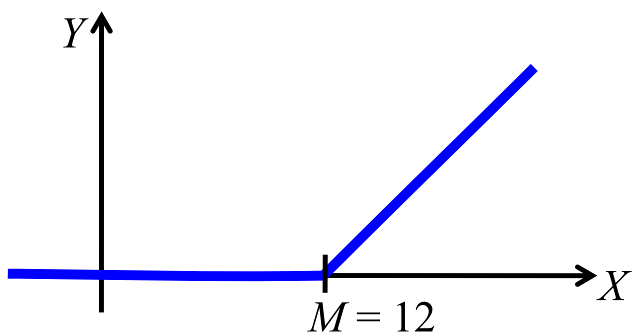
Let's start with

Discrete X, Discrete Y

Example 4.29 (**Discrete X, Discrete Y**)

- Let X be the number of active speakers in a group of $N = 48$ independent speakers, each with probability $p = 1/3$ of being active.
- A voice transmission system can transmit up to $M = 12$ voice signals at a time, and when X exceeds M, X-M signals are discarded.
- Let Y be the number of signals discarded.

$$Y = (X - M)^{+} \quad \text{where} \quad (x)^{+} = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$



Example 4.29 (Solution)

- The number of active speakers, X, is a binomial random variable with parameters N and p.

- Thus, X ranges over integer values from 0 to N and the number of missed packets, Y, ranges over 0 to N-M.

- The equivalent event for $\{Y = 0\}$ is $\{0 \le X \le M\}$. Thus, $P[Y=0] = \sum_{i=0}^{M} p_X(i)$

- The equivalent event for $\{Y = k\}$ is $\{X = M+k\}$ where $k \in \{1,\ldots, N\text{-}M\}$.

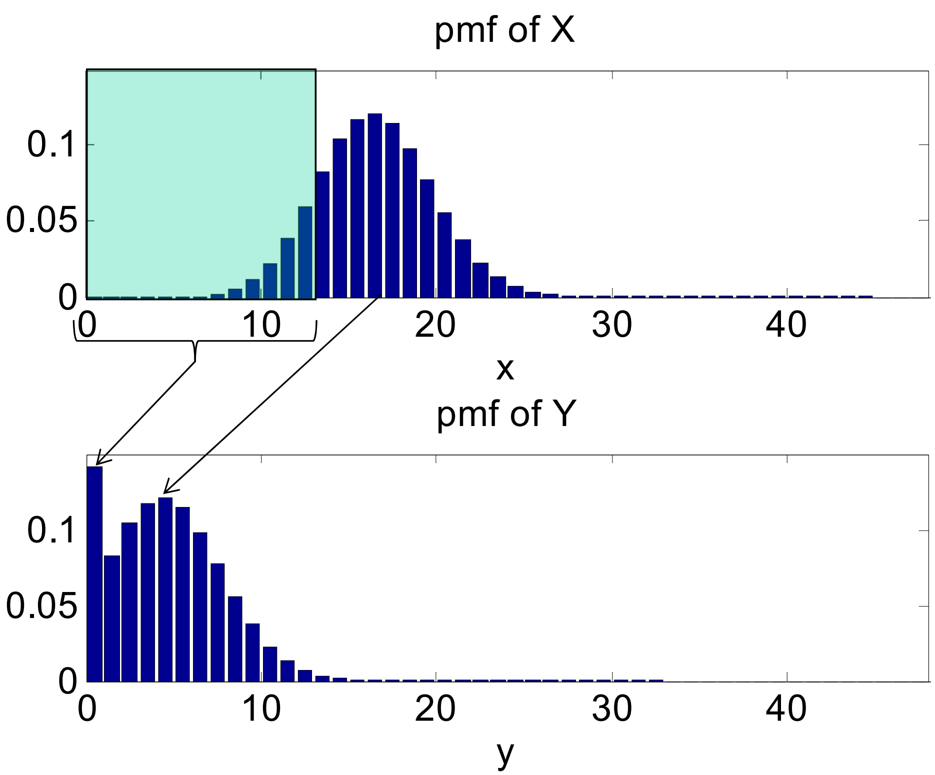
$$P[Y = k] = p_X(M+k) \text{ for } k \in \{1,\ldots,N-M\}$$

- Otherwise, $P[Y = k] = 0$.

Example 4.29 (Summary)

$$p_Y(k) = \begin{cases} \sum_{i=0}^{M} p_X(i) & \text{if } k = 0 \\ p_X(M+k) & \text{if } k \in \{1, \ldots, N-M\} \\ 0 & \text{otherwise} \end{cases}$$

$$p_X(i) = \binom{N}{i} p^i (1-p)^{N-i}$$
$$\text{for } i \in \{0, \ldots, N\}$$



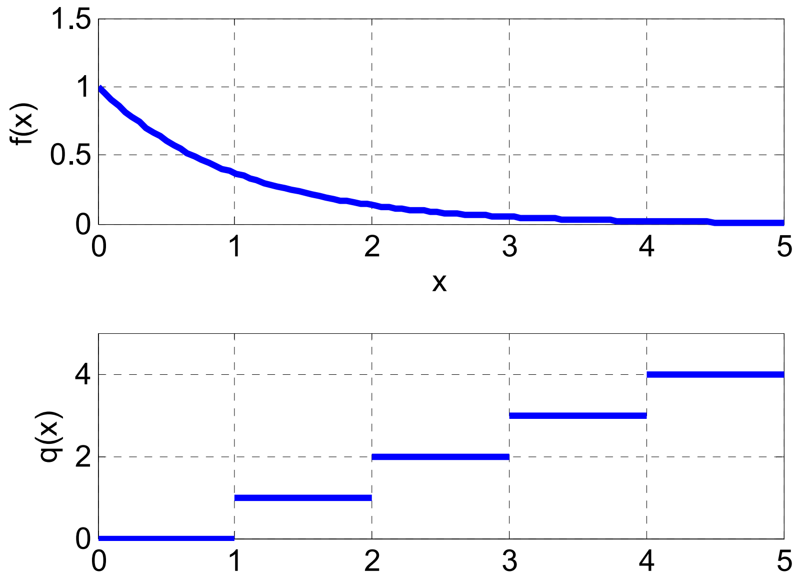
Now consider
Continuous X, Discrete Y

Example (**Continuous *X*, Discrete *Y***)

Suppose X is exponentially distributed with parameter λ = 1.

X is quantized by $q(x)$, which discards the digits past the decimal place.

Find the probability mass function for $Y = q(X)$.



$$f_X(x) = \lambda e^{-\lambda x} u(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$q(x) = \begin{cases} 0 & 0 \leq x < 1 \\ 1 & 1 \leq x < 2 \\ 2 & 2 \leq x < 3 \\ \vdots & \vdots \end{cases}$$

Solution

Y is discrete since it only takes values of $0,\ 1,\ 2,\ 3,\ldots.$

For $k = 0,1,2,3,\ldots.$

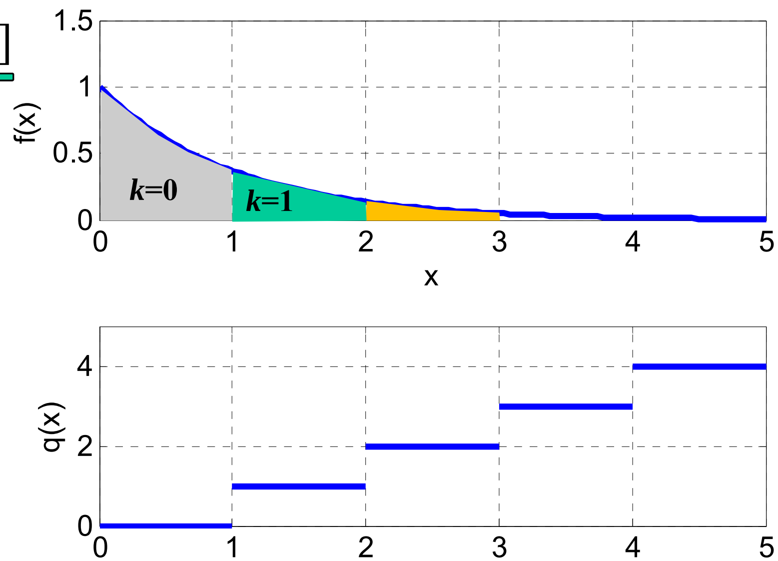
$$\underline{p_Y(k) = P[k \le X < k+1]}$$

$$= \underline{\int_k^{k+1} \lambda e^{-\lambda x}\,dx}$$

$$= -e^{-\lambda x}\Big|_k^{k+1}$$

$$= e^{-\lambda k} - e^{-\lambda(k+1)}$$

$$\boxed{= (1-e^{-\lambda})e^{-\lambda k}}$$



Otherwise, $p_Y(k) = 0$.

Hence, Y is a **geometric** RV with $p = (1\text{-}e^{-l})$: $p_Y(k) = (1-p)^k\, p$

Finally,
Continuous X, Continuous Y

Example 4.31: ***Linear Function*** of a Continuous RV

Let $g(x) = ax + b$ where $a, b \in \mathbb{R}$ and $a \neq 0$.

If X is a continuous RV with density $f_X(x)$, what is the density of $Y = g(X)$.

If $a > 0$:

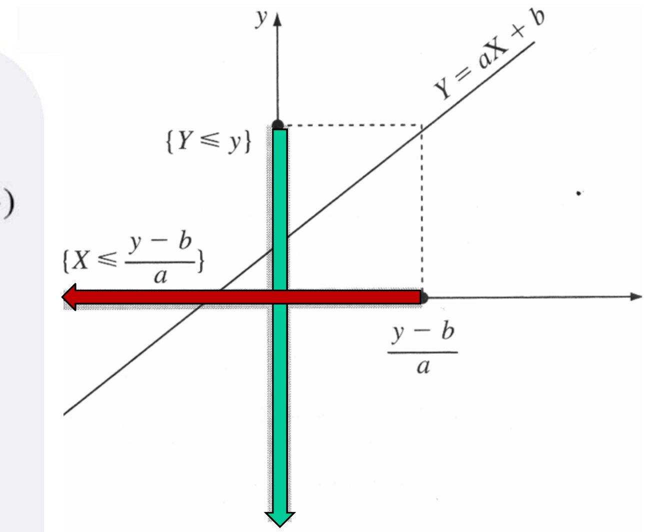
$$
\begin{aligned}
F_Y(y) &= P(\{Y \le y\}) \\
&= P(\{aX + b \le y\}) \\
&= P\left(\left\{X \le \frac{y-b}{a}\right\}\right) \\
&= F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \\
&= \boxed{\frac{1}{a} f_X\left(\frac{y-b}{a}\right)}
\end{aligned}
$$

If $a < 0$:

$$
\begin{aligned}
F_Y(y) &= P(\{Y \le y\}) \\
&= P(\{aX + b \le y\}) \\
&= P\left(\left\{X \ge \frac{y-b}{a}\right\}\right) \\
&= 1 - F_X\left(\frac{y-b}{a}\right)
\end{aligned}
$$

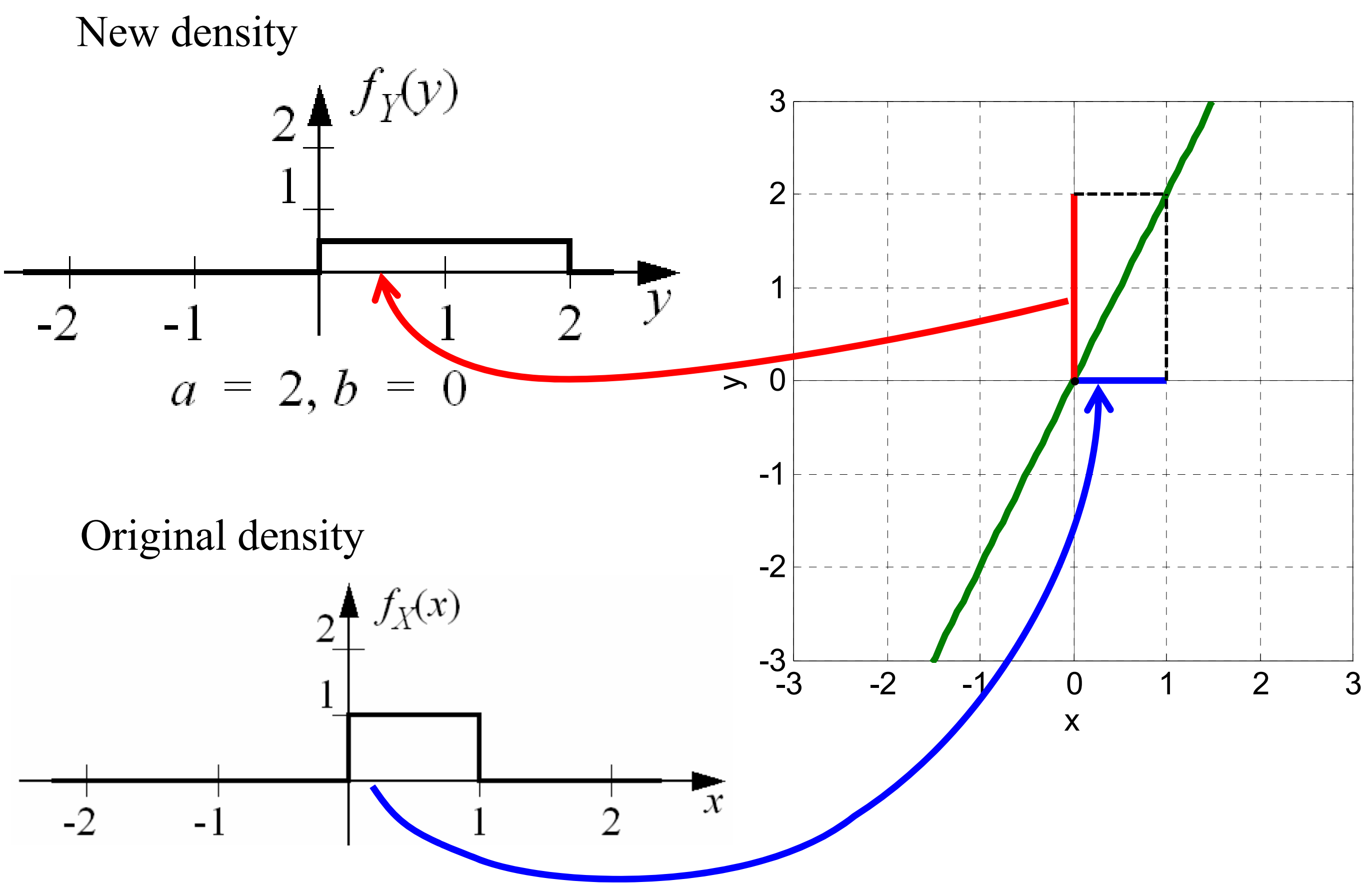
$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy}\left(1 - F_X\left(\frac{y-b}{a}\right)\right) \\
&= \boxed{\frac{1}{-a} f_X\left(\frac{y-b}{a}\right)}
\end{aligned}
$$



$$\boxed{\text{Thus, } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).}$$

Example: $a = 2,\ b = 0$

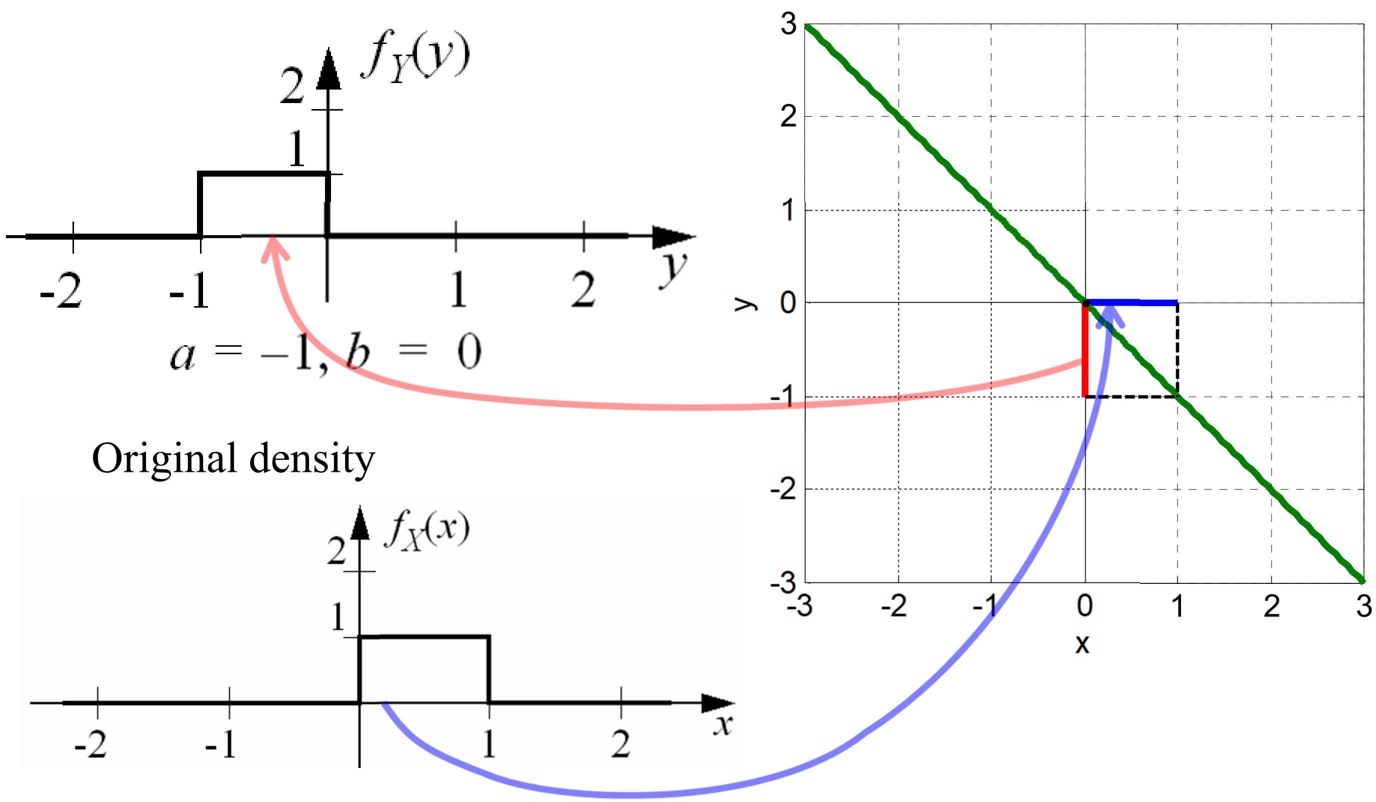
Suppose that X is uniformly distributed on [0,1]. Let $Y = aX + b$.



Example: $a = -1,\ b = 0$

Suppose that X is uniformly distributed on $[0,1]$. Let $Y = aX + b$.

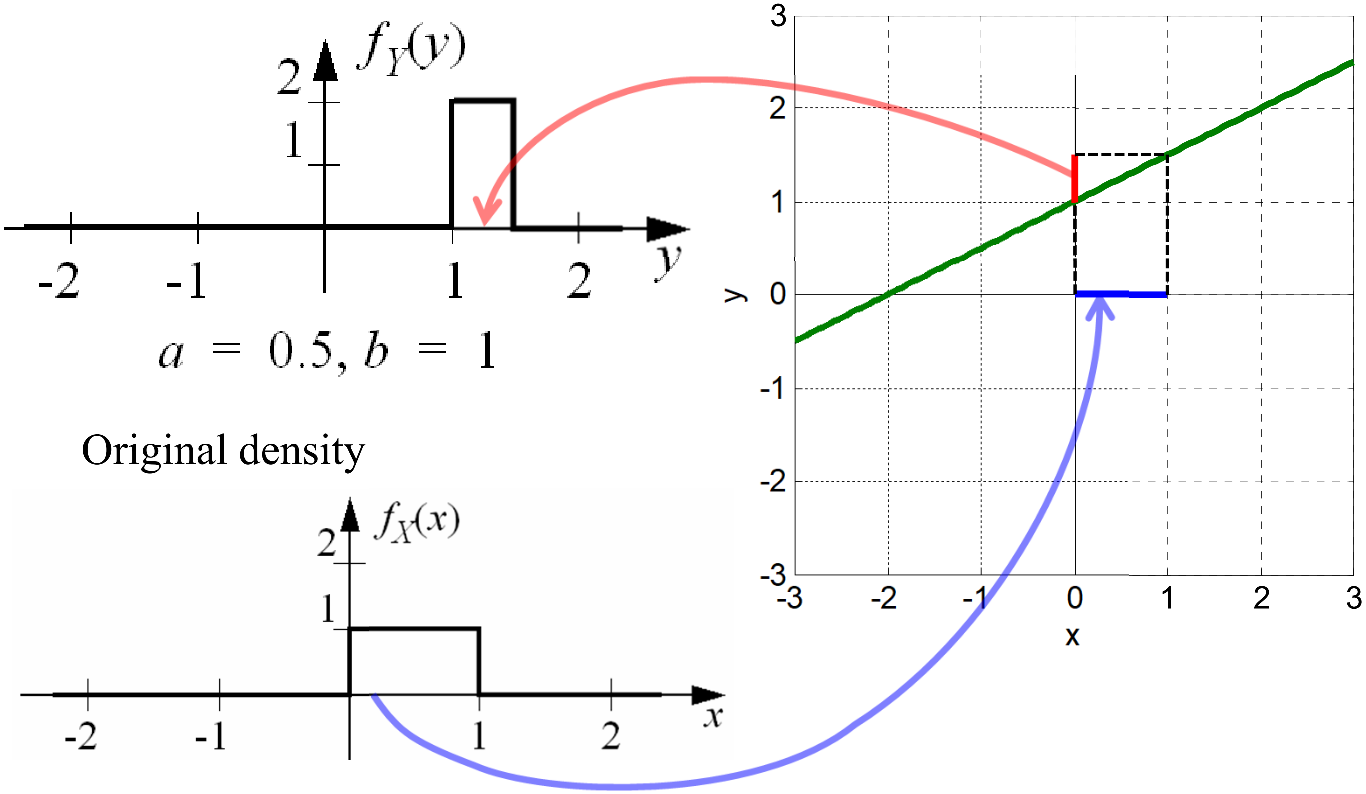
New density



Example: $a = 0.5$, $b = 1$

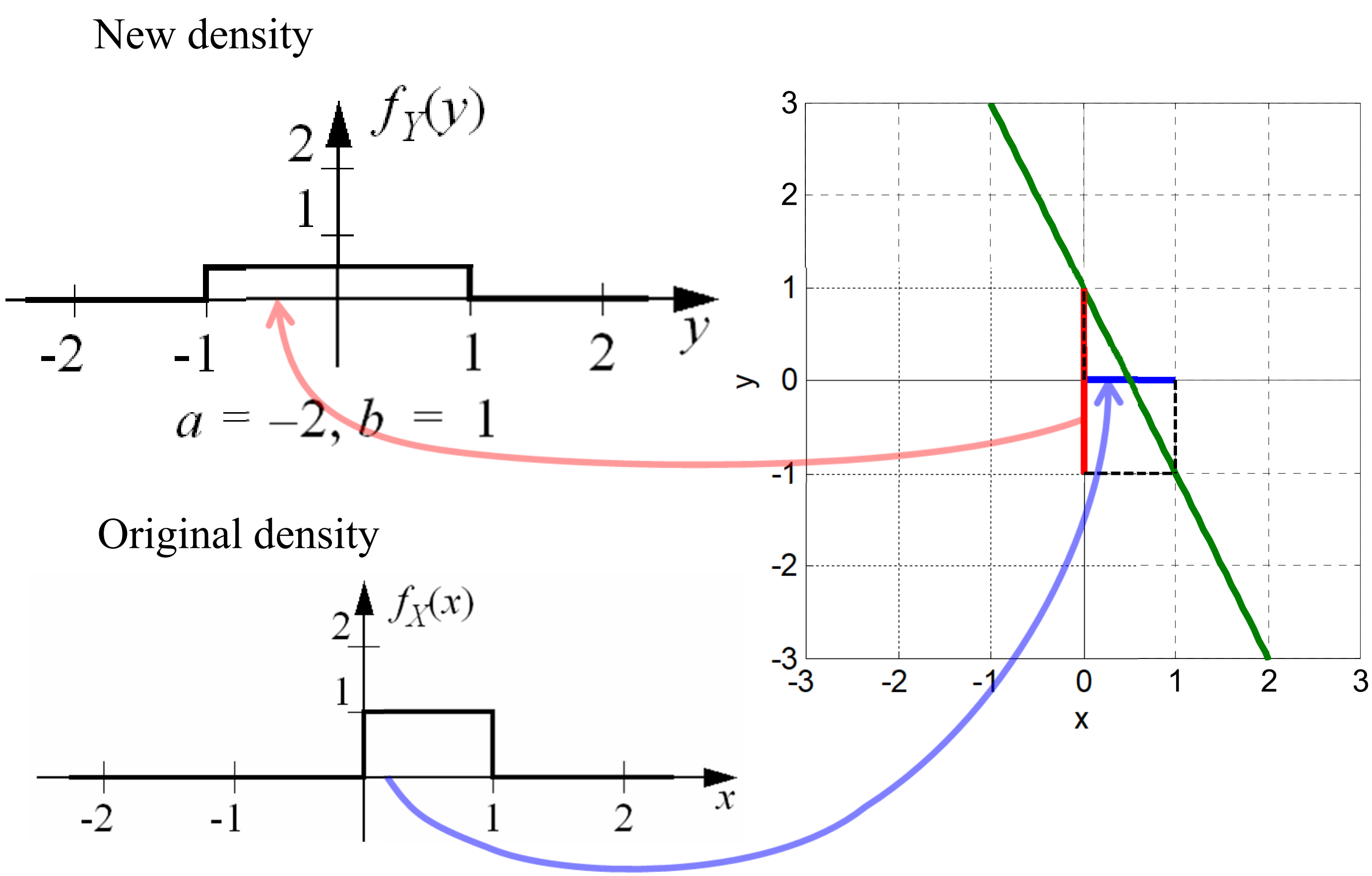
Suppose that X is uniformly distributed on $[0,1]$. Let $Y = aX + b$.

New density



Example: a = -2, b = 1

Suppose that X is uniformly distributed on $[0,1]$. Let $Y = aX + b$.



Example 4.32: Linear Function of a ***Gaussian*** RV

Let $g(x) = ax + b$ where $a, b \in \mathbb{R}$. Assume X is a Gaussian RV with density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

what is the density of $Y = g(X)$?

Solution:

Applying Example , $f_Y(y) = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{(y-b-am)^2}{2(a\sigma)^2}}$.

Y is also a Gaussian RV with mean $am + b$ and variance $(a\sigma)^2$.

Important Fact:

Any linear function of a Gaussian random variable is also Gaussian.

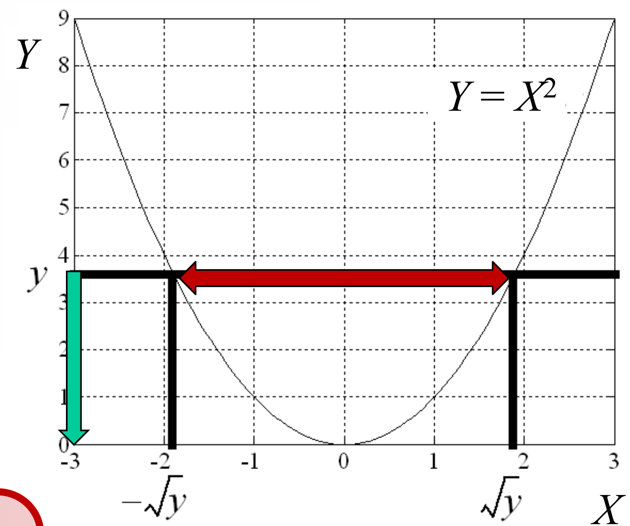
Example 4.33: Continuous X, Continuous Y

Suppose X is a continuous RV with density $f_X(x)$, what is the density of $Y = g(X)$ where $g(x) = x^2$.

Solution:

Find the cumulative distribution function

$$F_Y(y) = P(Y \le y) = \begin{cases} \displaystyle\int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\,dx & y > 0 \\ 0 & \text{otherwise} \end{cases}$$



Then differentiate,

$$f_Y(y) = \begin{cases} \dfrac{f_X(\sqrt{y})}{2\sqrt{y}} + \dfrac{f_X(-\sqrt{y})}{2\sqrt{y}} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Example 4.34: A Chi-Square Random Variable

Suppose that X is a ***Gaussian RV*** with zero mean and unit variance. Find the density of $Y = X^2$.

Solution:

From the previous example,

$$f_Y(y) = \begin{cases} \dfrac{f_X(\sqrt{y})}{2\sqrt{y}} + \dfrac{f_X(-\sqrt{y})}{2\sqrt{y}} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

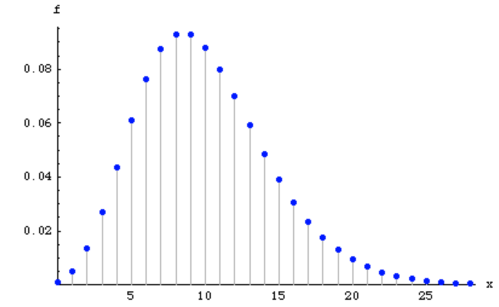
Substituting $f_X(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, we obtain

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{2y\pi}}$$

which is called the *chi-square random variable with one degree of freedom*.

Transform(er)s Ending

ELEC 2600: Probability and Random Processes in Engineering



Part II: Single Random Variables

- Lecture 5: Discrete Random Variables
- Lecture 6: Expected Value and Moments; Important Discrete Random Variables
- Lecture 7: Continuous Random Variables
- Lecture 8: Expected Value and Moments of Continuous Random Variables ; Important Continuous RVs
- Lecture 9: Function of a Random Variable
- **Out-of-Class Reading: Conditional PMF and Expectation**

Elec 2600H: Conditional PMF and Expectation

- **Conditional Probability Mass Function**

- Conditional Expected Value

Conditional Probability Mass Function

- The effect of partial information about the outcome of a random experiment on the probability of a discrete random variable is reflected by the ***conditional probability mass function.***
- Suppose that we know that an event C has occurred (we assume C has nonzero probability). The conditional probability mass function of X given C is

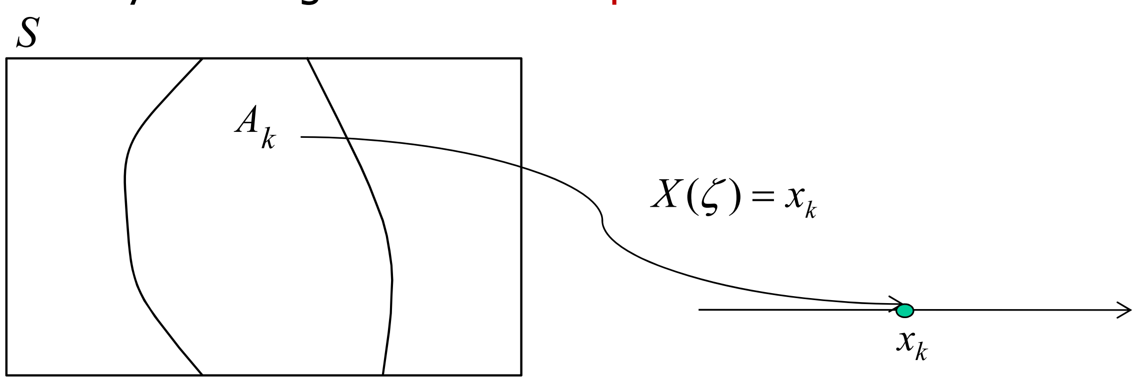
$$\boxed{p_X(x|C) = P[X = x|C]}$$

- By the definition of conditional probability

$$\boxed{p_X(x|C) = \frac{P[\{X = x\} \cap C]}{P[C]}}$$

Interpretation:

- Start by recalling unconditional pmf:



Event: $\zeta \in A_k$

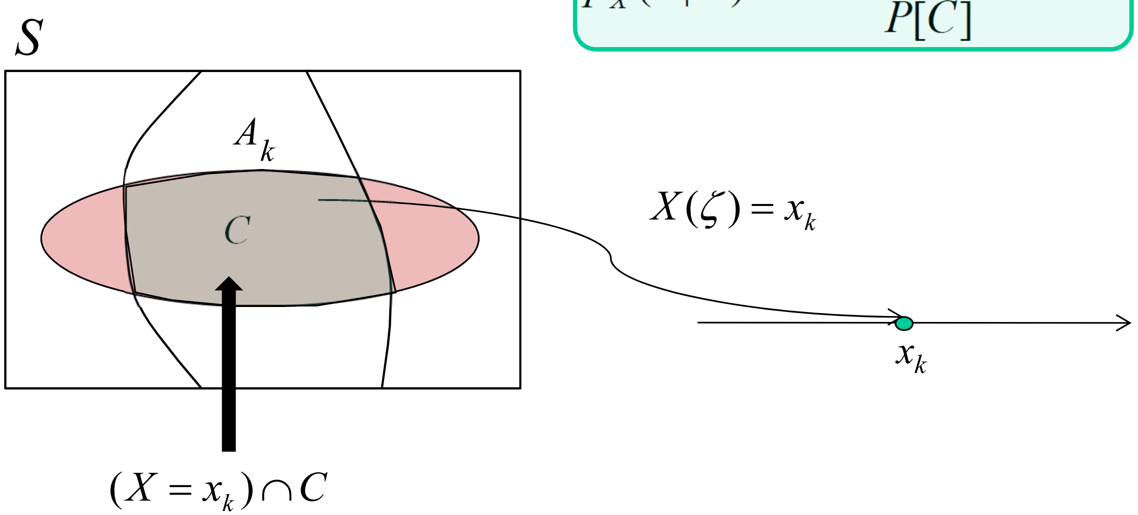
Equivalent Event

The pmf value $p_X(x_k) = P[X = x_k]$ is equivalent to $P[\zeta \in A_k]$

Interpretation:

- Now, with conditioning…

$$p_X(x \mid C) = \frac{P[\{X = x\} \cap C]}{P[C]}$$



Properties of the Conditional PMF

- The conditional pmf has the **same properties as the pmf**.

- The pmf is non-negative: $\boxed{p_X(x \mid C) \geq 0}$
- The values of the pmf sum to 1:

$$\boxed{\sum_{x \in S_X} p_X(x \mid C) = \sum_{\text{all } k} p_X(x_k \mid C) = 1}$$

- The conditional probability of events B defined by X can be computed by summing the conditional pmf:

$$\boxed{P\left[X \text{ in } B \mid C\right] = \sum_{x \in B} p_X\left(x \mid C\right) \text{ where } B \subset S_X}$$

Example: Waiting for Bus

- A person arrives at a bus stop at time ζ (minutes), taking values in $[1\ldots60]$ with equal probability
- There are 12 buses, with bus n arriving at time $5n$; i.e., the person takes bus 1 if arriving at time $\{1, 2, 3, 4, 5\}$, bus 2 if arriving at time $\{6, 7, 8, 9, 10\}$, etc.
- Let X be the number of the bus which the person takes. Find the conditional pmf of X **given** B = "the person takes one of the first four buses".

Example: Waiting for Bus

Solution

Since the arrival time is uniformly distributed and busses arrive regularly, the bus number is also uniformly distributed:

Since $B = [X \in \{1, 2, 3, 4\}]$,

$$p_X(k) = P[X = k] = P[\zeta \in \{5k-4, \ldots, 5k\}] = 5 \times \tfrac{1}{60} = \tfrac{1}{12} \quad \text{for } k = 1, 2, \ldots, 12$$

$$p_X(j|B) = \frac{P\big[\{X = j\} \cap \{X \in \{1,2,3,4\}\}\big]}{P[X \in \{1,2,3,4\}]} = \begin{cases} \dfrac{P[\{X = j\}]}{1/3} = \dfrac{1/12}{1/3} = \dfrac{1}{4} & if\ j \in \{1,2,3,4\} \\ 0 & otherwise \end{cases}$$

Example: Waiting for Bus

Let X be the number of the bus the person takes in the previous example. Find the conditional pmf of X **given**

D = "the person arrives during minute [2,...,11]"

Solution

Since $D = \{1 < z \leq 11\}$, $P[D] = P[\zeta \in \{2,...,11\}] = 10 \times \frac{1}{60} = \frac{10}{60}$

$$p_X(j|D) = \frac{P[\{X = j\} \cap D]}{P[D]} = \frac{P[\zeta : X(\zeta) = j \ and \ \zeta \in \{2, \ldots, 11\}]}{10/60}$$

$$= \begin{cases} \dfrac{P[\zeta \in \{2,3,4,5\}]}{10/60} = \dfrac{4}{10} & if \ j = 1 \\ \dfrac{P[\zeta \in \{6,7,8,9,10\}]}{10/60} = \dfrac{5}{10} & if \ j = 2 \\ \dfrac{P[\zeta \in \{11\}]}{10/60} = \dfrac{1}{10} & if \ j = 3 \\ 0 & otherwise \end{cases}$$

Conditioning Event *Defined by X*

- In many instances, the conditioning event, C, is defined in terms of X
  - E.g. $C = \text{"}X > 10\text{"}$ or $C = \text{"}a \le X \le b\text{"}$
- In this case, the conditional pmf can be determined **solely from the pmf of X**.

- For x_k in S_X, $$\boxed{p_X(x_k|C) = \frac{P[\{X=x_k\}\cap C]}{P[C]} = \begin{cases} \dfrac{p_X(x_k)}{\sum_{x\in C} p_X(x)} & if\, x_k \in C \\ 0 & if\, x_k \notin C \end{cases}}$$

From PMF
to conditional PMF

Example 3.24: Residual Waiting Times

- Let X, the time required to transmit a message, be a uniform random variable with $S_X = \{1, 2, \ldots, 20\}$: $p_X(k) = \frac{1}{20}$ for $k = 1, 2, \ldots, 20$

  Find the conditional pmf assuming that the message is *still not finished* transmitting at time 12.

Solution

Let C = "the message not finished at time 12" $= (X \in \{13, 14, \ldots, 20\})$. Since there are 8 outcomes in C and X is uniform, $P[C] = 8/20$.

$$p_X(j \mid C) = \frac{P[X = j \text{ and } X \in \{13,14,\ldots,20\}]}{P[C]}$$

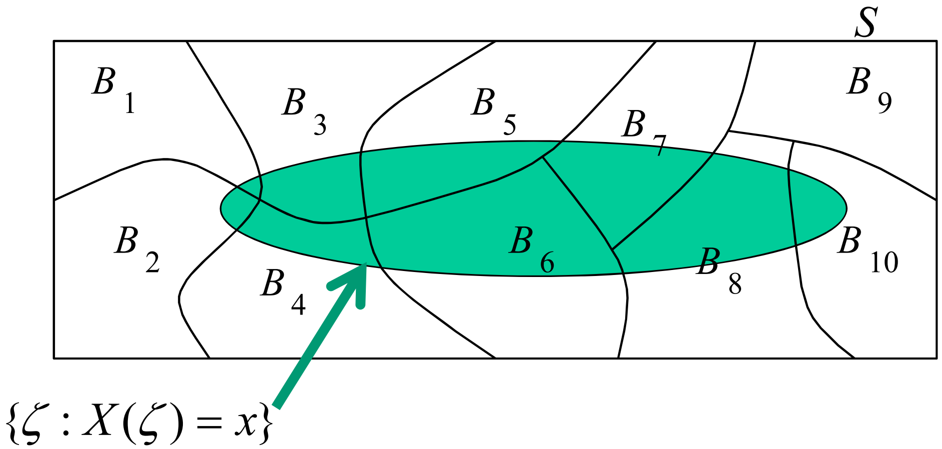
$$= \begin{cases} \dfrac{p_X(j)}{8/20} = \dfrac{1/20}{8/20} = \dfrac{1}{8} & \text{if } j \in \{13,14,\ldots,20\} \\ 0 & \text{otherwise} \end{cases}$$

Total Probability Theorem

- In many cases, it may be natural to analyze the outcome of an experiment by conditioning on individual events $B_1, B_2, \ldots B_n$ in a partition of a sample space.
- The **theorem on total probability** allows us to combine conditional pmfs to find the pmf of X

$$p_X(x) = \sum_i p_X(x|B_i)P[B_i]$$

What's the difference from the previous one?



Example 3.25: Device Lifetimes

- A production line yields two types of devices
  - **Type 1**: *short* lifetimes geometrically distributed with parameter r
  - **Type 2**: *long* lifetimes geometrically distributed with parameter s
- Suppose the devices occur with probabilities α (Type 1) and (1-α) (Type 2).
- Find the pmf of the lifetime of an **arbitrary device**, X.

Solution

Let B_1 = "device is Type 1" and B_2 = "device is Type 2".

Since

$$p_X(k \mid B_1) = (1-r)^{k-1} r \qquad for \qquad k = 1,2,3\ldots$$

$$p_X(k \mid B_2) = (1-s)^{k-1} s \qquad for \qquad k = 1,2,3\ldots$$

$$P[B_1] = \alpha$$

$$P[B_2] = 1-\alpha$$

We have that

$$p_X(k) = p_X(k \mid B_1)P[B_1] + p_X(k \mid B_2)P[B_2]$$

$$= (1-r)^{k-1} r\alpha + (1-s)^{k-1} s(1-\alpha) \quad \text{for } k = 1, 2, 3, \ldots.$$

Elec 2600H: Conditional PMF and Expectation

- Conditional Probability Mass Function
- **Conditional Expected Value**

Conditional Expected Value and Variance

- The conditional pmf has *all the properties of a regular pmf*. Thus, we can define the same quantities on the conditional pmf as we did for the regular pmf, including **expectation**.
- **Definition: The conditional expected value of X given B is**

$$m_{X|B} = E[X|B] = \sum_{x \in S_X} x p_X(x|B) = \sum_k x_k p_X(x_k|B)$$

- Interpretation: The average of X in a large number of trials ***where B occurs***
- **Definition: The conditional variance of X given B is**

$$\begin{aligned} \mathrm{VAR}[X|B] &= E[(X - m_{X|B})^2 | B] = \sum_k (x_k - m_{X|B})^2 p_X(x_k|B) \\ &= E[X^2|B] - m_{X|B}^2 \end{aligned}$$

- Interpretation: The average squared distance of X from its mean over a large number of trials ***where B occurs***.

Using Conditioning to Compute Expectations

- In cases where it is natural to analyze the outcome of an experiment by *conditioning on individual events* $B_1, B_2, \ldots B_n$ in a partition of a sample space, we can **compute expected values from conditional expected values**:

$$\boxed{E[X] = \sum_i E[X \mid B_i] P[B_i]}$$

Proof

$$E[X] = \sum_k x_k p_X(x_k)$$

$$= \sum_k x_k \left\{ \boxed{\sum_i p_X(x_k \mid B_i) P[B_i]} \right\}$$

$$= \sum_i \left\{ \boxed{\sum_k x_k p_X(x_k \mid B_i)} \right\} P[B_i]$$

$$= \sum_i E[X \mid B_i] P[B_i]$$

Example 3.25a: EV of Device Lifetimes

- A production line yields two types of devices
  - o **Type 1**: short lifetimes geometrically distributed with parameter r
  - o **Type 2**: long lifetimes geometrically distributed with parameter s
- Suppose the devices occur with probabilities α (Type 1) and ($1-\alpha$) (Type 2).
- Find the expected value of the lifetime of an arbitrary device, X.

Solution

Let B_1 = "device is Type 1" and B_2 = "device is Type 2".

Since $m_{X|B_1} = 1/r \qquad P[B_1] = \alpha$

$m_{X|B_2} = 1/s \qquad P[B_2] = (1-\alpha)$

We have that
$$\begin{aligned} m_X &= m_{X|B_1} \cdot P[B_1] + m_{X|B_2} \cdot P[B_2] \\ &= 1/r \cdot \alpha + 1/s \cdot (1-\alpha) \\ &= \alpha/r + (1-\alpha)/s \end{aligned}$$

Example 3.25b: Variance of Device Lifetimes

- Find the variance of X, the lifetime of an arbitrary device, in the previous example

Solution

Let B_1 = "device is Type 1" and B_2 = "device is Type 2".

It can be shown that the variance of a geometric random variable with parameter p is given by $(1-p)/p^2$

Thus, $$\mathrm{E}[X^2 \mid B_1] = \mathrm{Var}[X \mid B_1] + m_{X|B_1}^2 = \frac{1-r}{r^2} + \frac{1}{r^2} = \frac{2-r}{r^2}$$

$$\mathrm{E}[X^2 \mid B_2] = \frac{2-s}{s^2}$$

Example 3.25b (cont)

Combining these by total probability, we obtain

$$E[X^2] = E[X^2 \mid B_1] \cdot P[B_1] + E[X^2 \mid B_2] \cdot P[B_2]$$
$$= \frac{2-r}{r^2} \cdot \alpha + \frac{2-s}{s^2} \cdot (1-\alpha)$$

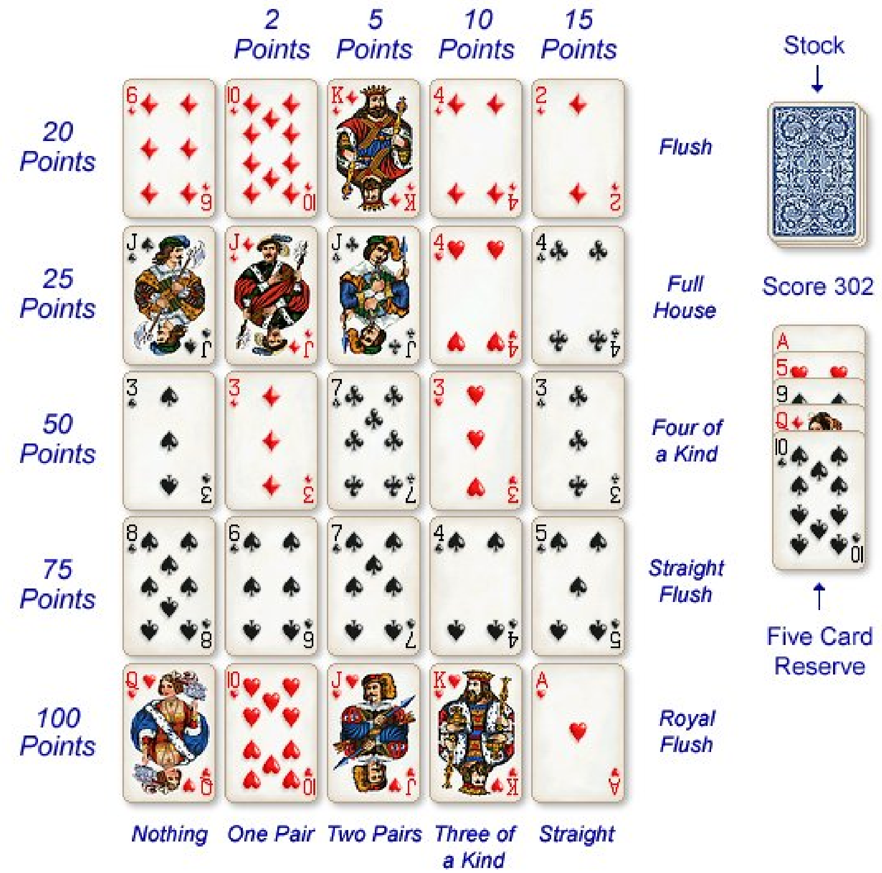
Thus,

$$\mathrm{VAR}[X] = E[X^2] - m_X^2$$
$$= \frac{2-r}{r^2} \cdot \alpha + \frac{2-s}{s^2} \cdot (1-\alpha) - \left(\frac{\alpha}{r} + \frac{1-\alpha}{s}\right)^2$$

Note that

$$\mathrm{VAR}[X] \neq \mathrm{VAR}[X \mid B_1] \cdot P[B_1] + \mathrm{VAR}[X \mid B_2] \cdot P[B_2]$$
$$\neq \frac{1-r}{r^2} \cdot \alpha + \frac{1-s}{s^2} \cdot (1-\alpha)$$

Poker Time



2 Points
5 Points
10 Points
15 Points
Stock
20 Points
Flush
Score 302
25 Points
Full House
50 Points
Four of a Kind
75 Points
Straight Flush
Five Card Reserve
100 Points
Royal Flush
Nothing
One Pair
Two Pairs
Three of a Kind
Straight

Straight Flush!

Oh Yeah!

Chance to Win?

- What is the chance that the cat can beat the president?
- With what probability, for each poker hand, can the president win?
- Conditional PMF?