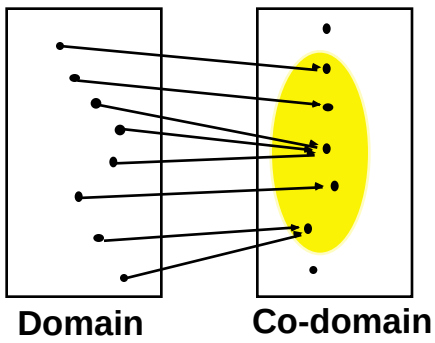
Chapter 4 Random Variables

- Random variables are used to model phenomena in which the experimental outcomes are *numbers*, e.g. 1, 2, 3, or 3.213678… instead of labels such as Head or Tail or Luke or Darth
- Example: $\Omega =$, or $\Omega = Z$
- We do not know for sure which number will be observed when the experiment is performed; only that it is some number in the sample space Ω
- **X** denotes the random number that we observe. It is called a *random variable.*
- A different number (value of **X**) is typically observed on each trial of the experiment
- Hence – *variable*
- Number **X** is an outcome of a random experiment
- Hence – *random*
- Alternative formulation
- A random variable **X** is obtained by associating real numbers with the outcomes of random experiments
- We have already used this idea before, but we did not bother to give it a formal name
- **Example:** The experiment consists of tossing a coin twice. **X** denotes the number of heads observed

Outcome	Value of **X**
HH	2
HT	1
TH	1
TT	0

- In this alternative formulation, **X** is *also* thought of as the *function* or *mapping* that maps Ω into real numbers
- Digression for brief review
- A *function* is a mapping from a set called the *domain* to a set called the *co-domain*
- A function f with domain A and co-domain B is denoted $f: A \rightarrow B$
- Every x A is associated with some y B
- y is the image of x, and x is the pre-image of y. We write $f(x) = y$
- Every x A has an image f(x) B
- Every y B need not have a pre-image in A
- An element y B might have more than one pre-image in A
- The set of all y B that have pre-images in A is called the *range* of f
- In the picture shown below, the range is the shaded oval-shaped set.



- A random variable $\mathbf{X}$ is a function $\mathbf{X}: \Omega \rightarrow$. Here, denotes the set of all real numbers
- Every outcome in the sample space Ω is mapped onto a real number by the function $\mathbf{X}$
- The real number assigned to ω Ω is denoted by $\mathbf{X}(\omega)$
- Example:

ω	$\mathbf{X}(\omega)$
HH	2
HT	1
TH	1
TT	0

- There is *nothing random* about this function. It *always* maps HH onto 2, for example
- The randomness arises from the fact that we do not know which of the 4 outcomes will occur, and hence we do not know which of the 3 numbers 0, 1, 2 will be observed
- The numbers 0, 1, 2 occur with probabilities q^2, 2pq and p^2 respectively; p = P(H) = 1–q
- In this alternative formulation, the function $\mathbf{X}: \Omega \rightarrow$ is a fixed (i.e., nonrandom) map
- ω Ω always has the *same real number* $\mathbf{X}(\omega)$ as its image
- Randomness lies in which outcome ω Ω occurred, and not in the mapping – the mapping is unchanging
- The observed value $\mathbf{X}(\omega)$ depends on which ω occurred on the trial of the experiment
- **Nomenclature:** Although $\mathbf{X}$ is a function $\Omega \rightarrow$, we look at the random values observed on repeated trials, and conclude that we are dealing with a random variable
- **Nomenclature:** Although $\mathbf{X}$ is a function $\Omega \rightarrow$, the observed values $\mathbf{X}(\omega)$ are a random variable
- More nomenclature: We drop the explicit dependence on ω and just write $\mathbf{X}$ instead of $\mathbf{X}(\omega)$
- **Mini-digression:** Dropping the argument of a function is a despicable trick commonly used by mathematicians and engineering professors to confuse engineering students
- If u(x) and v(x) are functions of x, then $\frac{d\ u/v}{dx} = \frac{v\ du - u\ dv}{v^2}$
- $\frac{d}{dx}\frac{HI}{LO} = ?$

  LO d(HI) minus HI d(LO), E-I-E-I-O
  Underneath, there's the square of LO, E-I-E-I-O
  With a dx here, and a dy there,
  Yes, there's hope you can cope
  Oh, the joys of calculus, E-I-E-I-O
- If x(t) is the input to a linear time-invariant system with impulse response h(t), then the output y(t) is given by

$$y = x*h = h*x$$

- **End of mini-digression**
- The function $\mathbf{X}: \Omega \rightarrow$ gives us real numbers to associate with outcomes
- There is nothing random about the mapping — it always assigns the same number to each outcome
- The values $\mathbf{X}(\omega)$ that we observe over a sequence of repeated trials change randomly from trial to trial as different outcomes occur

- We abuse notation and say that we have a random variable $\mathbf{X}$ (instead of $\mathbf{X}(\omega)$)
- We drop the explicit dependence on ω when we are talking in general about the random values observed
- We shall say "Let $\mathbf{X}$ denote a random variable …"
- We will sometimes need to discuss a specific outcome ω_1 and its image, which is the real number $\mathbf{X}(\omega_1)$
- Why bother with the alternative approach with mappings etc. at all? Why not just think of random variables as real numbers that we observe directly?
- In many instances, the notion of associating real numbers with actual outcomes is more natural or reasonable
- **Example:** An experiment is repeated 4 times. Let the random variable $\mathbf{X}$ denote the number of times that an event A occurred on these 4 trials. Thus, $\mathbf{X}(A, A, A^c, A^c) = 2$ and $P(\mathbf{X} = 2) = \binom{4}{2} p^2 q^2$
- Fundamental scientific advances have been made by insisting that every random variable is a mapping from a suitably defined sample space to the real numbers
- The trick lies in being smart enough to figure out the sample space and the map
- **Example:** Pressure exerted by a gas is caused by the gas molecules striking the walls of the container. A suitable probabilistic model for the velocities of the molecules led to the Maxwell-Boltzmann theory and the development of statistical thermodynamics
- **Example:** By postulating a probabilistic model for electron energies in solids, we get important information about semiconductors and their properties, charge transport, properties and behavior of junctions, noise phenomena, etc
- **Example:** Simple ionization models lead to explanations of how and why RF signals seem to repeatedly fade away and then grow stronger again
- **Example:** Models of arrival processes allow us to study queues and find optimum buffer sizes etc
- We may not have a good model for the underlying sample space
- We may not understand the map from the sample space to the real numbers very well
- Understanding in detail might be left to future generations
- **Example:** Noise in a resistor is caused by the random motion of electrons. The exact voltage value is a function of all the positions of all the electrons. Maxwell's equations give the answer in theory, but the computational task is formidable
- We shall insist that a random variable always is a mapping from a sample space Ω to
  - — even if we do not have a good idea about the sample space
  - — even if we do not have a good idea about the exact mapping $\mathbf{X}: \Omega \rightarrow$ or cannot compute it exactly in a particular case
- **Summary:** A random variable $\mathbf{X}$ denotes two different things
- $\mathbf{X}$ is the name of the fixed, unchanging, nonrandom, function or rule that assigns real numbers to the outcomes in a sample space
- $\mathbf{X}: \Omega \rightarrow$
- Every outcome ω Ω is mapped onto the real number $\mathbf{X}(\omega)$
- Every real number need not have a pre-image in Ω
- More than one outcome may be mapped onto the same number

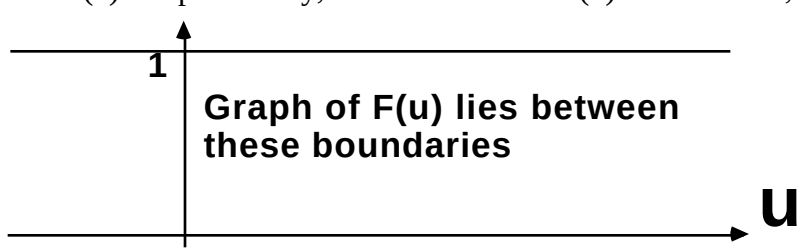
- **X** is also the name assigned to the numbers we observe when the experiment is repeated
- If the outcome on a particular trial is ω_1 Ω, we say that the random variable **X** has taken on value $\mathbf{X}(\omega_1)$ on this trial
- The values observed *vary randomly* as the experiment is the experiment is repeated — hence *random variable* **X**
- The *function* $\mathbf{X}: \Omega \rightarrow$ is unchanging — we always map HH onto 2 — but the observed value of **X** depends on which outcome occurred
- As trials of the experiment are performed, we can keep track of the observed values

- If a discrete set of values occurs repeatedly, our picture might look as follows

- A *discrete* random variable **X** takes on discrete values $u_1, u_2, u_3, \ldots u_n, \ldots$ where $u_1 < u_2 < u_3 < \ldots < u_n < \ldots$
- The range of **X** might be finite (n values $u_1, u_2, u_3, \ldots u_n$) or it might be countably infinite ($u_1, u_2, u_3, \ldots u_n, \ldots$)
- As a practical matter, we can never observe an infinite number of values, so some modeling is necessary
- Another possibility for the experiment is that different values occur on each and every trial

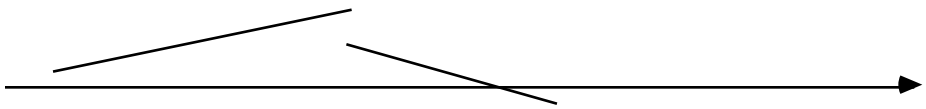
- If our line looks like this after a few billion trials, we say that **X** is a continuous random variable
- Once again, some modeling is involved in making the jump from the finite set of observed values to the continuum of values
- If the values taken on by **X** are different on each and every trial, we *assume* that **X** can take on any value in , or any value in an *interval* in
- We say that **X** is a *continuous* random variable with range or range (a,b) or (a,b) (c,d)…
- Assumption: **X** can take on any value in some interval Some modeling is involved in making the jump from the *finite set* of observed values to the *continuum* of values especially in jumping to the conclusion that **X** can be *any* real number
- **Example:** All the observed values of **X** are in the range (–219.82, +422.19)
- Should we model **X** as taking on all values in ? or just all values in the interval (–219.82, +422.19)? or just all values in the interval (–500,+500)?
- A *mixed* random variable takes some values repeatedly (just like a discrete random variable), while some other values are all scattered (and different) in an interval (like a continuous random variable)

- A random variable takes on different values as the experiment is repeated
- We want to describe this in probabilistic terms
- What is the probability that **X** takes on the value 9.621354?
- What is the probability that **X** ≤ 9.621354? or **X** ≤ 39?
- Probabilities represent relative frequencies
- If **X** takes on value 2 with probability 1/3, then 2 should occur on roughly one-third of the trials
- If **X** ≤ 0 with probability 1/2, positive numbers should occur on half the trials
- Simple cases lead to simple answers
- Let **X** be a discrete random variable where the underlying sample space and the map are well-understood
- Then, we can just compute probabilities directly
- **Example:** **X** denotes the number of heads observed on N independent tosses of a biased coin with P(H) = p
- **X** takes on values 0, 1, 2,…, N • $P\{\mathbf{X} = k\} = P(B(k;N)) = \binom{N}{k} p^k q^{N-k}$
- **X** is called a binomial random variable with parameters (N, p)
- **Example:** **X** denotes the number of independent tosses of a biased coin (with P(H) = p) needed to observe the first head
- **X** takes on values 1, 2, … • $P\{\mathbf{X} = k\} = P(C(k)) = pq^{k-1}$;
- **X** is called a geometric random variable with parameter p
- **Example:** **X** denotes the number of tosses of a biased coin (with P(H) = p) needed to observe the r-th head
- **X** takes on values r, r+1, … • $P\{\mathbf{X} = k\} = P(C(k;r)) = \binom{k-1}{r-1} p^r q^{k-r}$;
- **X** is a negative binomial random variable with parameters (r, p)
- Common approach (useable for defining probabilities for *all* random variables) is based on the cumulative probability distribution function (CDF or cdf)
- CDF is denoted F(u) or $F_{\mathbf{X}}(u)$
- u is any number, $-\infty < u < \infty$
- $F(u) = P\{\mathbf{X} \le u\}$ for all u
- The text uses the notation $F(x) = P\{\mathbf{X} \le x\}$
- Unfortunately, this leads to much confusion in students' minds
- What is the difference between **X** and x? • Is x the random variable?
- Since the value of F(u) is a probability, we know that $0 \le F(u) \le 1$ for all u, $-\infty < u < \infty$

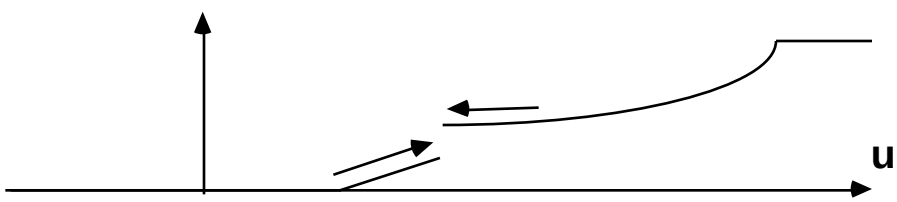


- $F(8.23) = P\{\mathbf{X} \le 8.23\}$ • $F(-38.76) = P\{\mathbf{X} \le -38.76\}$

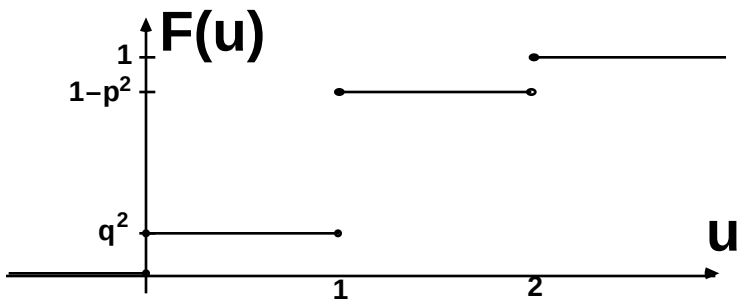
- Clearly, if we observe that **X** has value $-50.12 \le -38.76$, then this value is also ≤ 8.23
- $F(-38.76) \le F(8.23)$
- More generally, if α and β are real numbers with $\alpha < \beta$, then $F(\alpha) \le F(\beta)$
- $F(u)$ is a non-decreasing function of u
- $F(-\infty) = \lim_{u \to -\infty} F(u) = 0$
- $F(+\infty) = \lim_{u \to +\infty} F(u) = 1$
- $F(u) = \lim_{\delta \to 0, \delta > 0} F(u+\delta) = F(u^+)$
- These are examples of the set-continuity properties of probability:
  the probability of a limit is the limit of the probability
- Digression: f(x) is *continuous* at x = a if both $f(a+\delta)$ and $f(a-\delta)$ converge to the same limit f(a) as $\delta \to 0$
- f(x) has a *jump discontinuity* at x = a if both $f(a+\delta)$ and $f(a-\delta)$ converge as $\delta \to 0$ but only one of these limits converges to f(a)



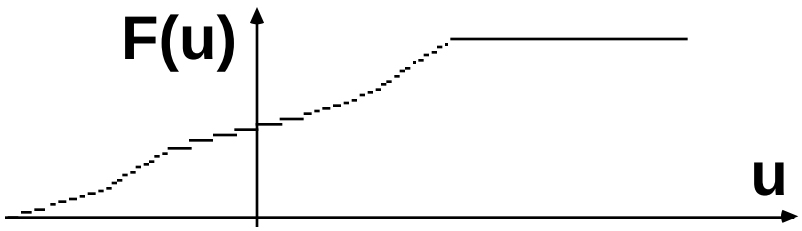
- $F(u) = \lim_{\delta \to 0, \delta > 0} F(u+\delta) = F(u^+)$
- This is saying that if F(u) has a jump discontinuity, then its value at the discontinuity is the limit as we approach from the right, not from the left



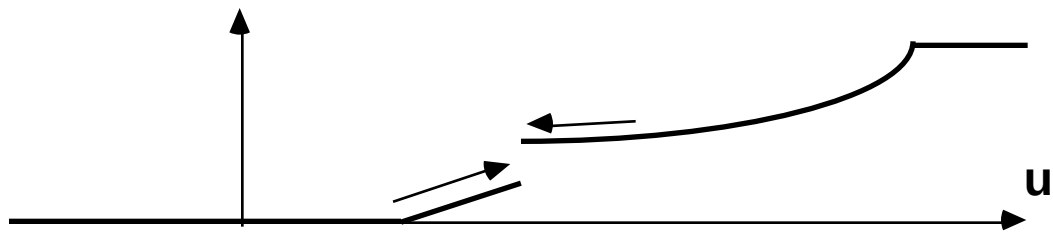
- F(u) is said to be *continuous from the right*, or a *right-continuous* function
- F(u) is either continuous, or if it has a jump discontinuity, then its value is the limit from the right (which is also the "upper" value)
- **Example:** **X** is the number of heads when a biased coin with P(H) = p is tossed twice
- **X** takes on values 0, 1, 2 with probabilities q^2, 2pq and p^2
- $P\{\mathbf{X} \le u\} = 0$ for any $u < 0$
- $P\{\mathbf{X} \le 0\} = q^2$
- **X** takes on values 0, 1, 2 with probabilities q^2, 2pq and p^2
- $P\{\mathbf{X} \le u\} = q^2$ for $0 \le u < 1$
- $P\{\mathbf{X} \le 1\} = q^2 + 2pq = 1-p^2$
- $P\{\mathbf{X} \le u\} = 1-p^2$ for $1 \le u < 2$
- $P\{\mathbf{X} \le 2\} = 1$
- $P\{\mathbf{X} \le u\} = 1$ for $2 \le u$
- $F(u) = \begin{cases} 0, & \text{for } u < 0 \\ q^2 & \text{for } 0 \le u < 1 \\ 1-p^2 & \text{for } 1 \le u < 2 \\ 1 & \text{for } 2 \le u \end{cases}$



- If Ω or the mapping $\mathbf{X}$ is not clearly defined and we only have observed values of $\mathbf{X}$ over many trials, the relative frequencies are used as the probabilities
- **Example:** $\mathbf{X}$ = 0,1,2 on 20, 62, 18 trials. The "jumps" are 0.2, 0.62, and 0.18 respectively
- **Example:** If we model $\mathbf{X}$ as a continuous random variable after observing the result of N trials, there are N jumps each of size 1/N. Approximate by a smooth function



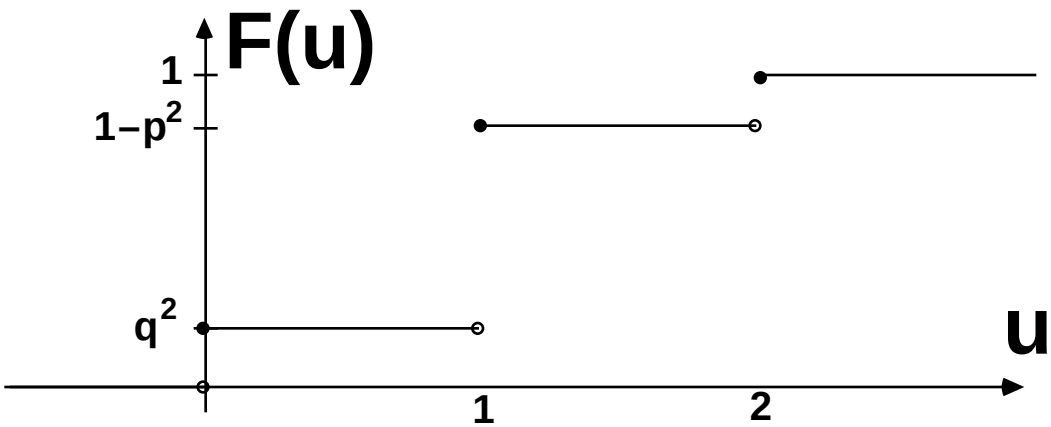
- The CDF includes all the probabilistic information about the random variable $\mathbf{X}$
- $P\{\mathbf{X} \le u\} = F(u) = F(u^+)$ • $P\{\mathbf{X} > u\} = 1 - F(u)$
- $P\{\mathbf{X} < u\} = F(u^-)$ = limit from the left as we approach u
- $P\{\mathbf{X} \le u\} = F(u) = F(u^+)$ • $P\{\mathbf{X} < u\} = F(u^-)$
- If F(u) is continuous at u, then $F(u^+) = F(u^-)$
- $\{\mathbf{X} \le u\} = \{\mathbf{X} < u\} \quad \{\mathbf{X} = u\}$
- If F(u) is continuous at u, then $P\{\mathbf{X} = u\} = 0$
- $P\{\mathbf{X} = u\} = 0$ *does not mean* that the event $\{\mathbf{X} = u\}$ will never occur. If it does occur at all, it will occur very infrequently and its relative frequency will converge to 0
- $\{\mathbf{X} \le b\} = \{\mathbf{X} \le a\} \quad \{a < \mathbf{X} \le b\}$ • $P\{a < \mathbf{X} \le b\} = F(b) - F(a)$
- CDF of a random variable $\mathbf{X}$ is denoted by F(u) or $F_{\mathbf{X}}(u)$
- u is any number, $-\infty < u < \infty$
- $F(u) = P\{\mathbf{X} \le u\}$ for all u
- F(u) is a *non-decreasing right-continuous* function with limiting values 0 at $-\infty$ and 1 at $+\infty$
- Nondecreasing: If $\alpha < \beta$, then $F(\alpha) \le F(\beta)$
- Right-continuous: $F(u) = \lim_{\delta \to 0, \delta > 0} F(u+\delta) = F(u^+)$
- Limiting values: $F(-\infty) = \lim_{u \to -\infty} F(u) = 0$ $F(+\infty) = \lim_{u \to +\infty} F(u) = 1$
- Note that this implies that $0 \le F(u) \le 1$
- At a jump discontinuity, the value of F is the limit as we approach from the right, and not the limit from the left



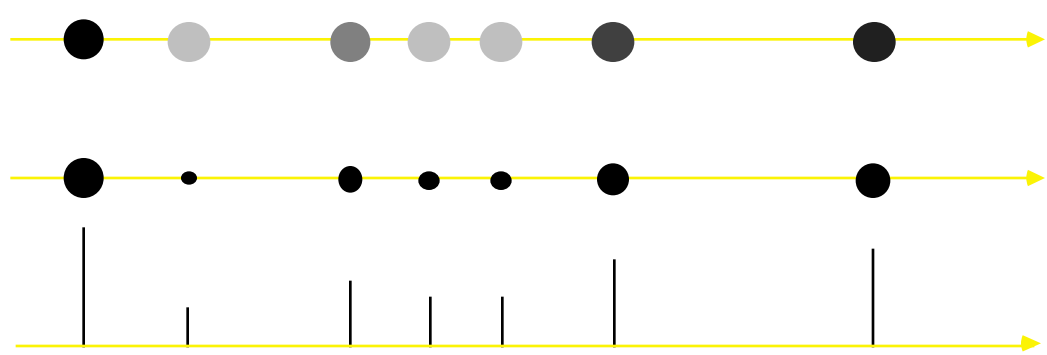
- At any number u, F(u) is either continuous, or it has a jump discontinuity. The value of F at the point u is the limit from the right (which is also the "upper" value in the case of a jump discontinuity)
- $F(u^-)$ = "lower value" = $\lim_{\delta \to 0, \delta > 0} F(u-\delta) = P\{\mathbf{X} < u\}$
- $P\{a < \mathbf{X} \le b\} = F(b) - F(a)$
- $P\{a < \mathbf{X} \le b\} = F(b^+) - F(a^+)$
- $P\{a \le \mathbf{X} \le b\} = F(b^+) - F(a^-)$
- $P\{a \le \mathbf{X} < b\} = F(b^-) - F(a^-)$
- $P\{a < \mathbf{X} < b\} = F(b^-) - F(a^+)$
- If F(u) is continuous, these formulas give same results
- If F(u) is continuous at a, then $P\{a < \mathbf{X} \le b\}=P\{a \le \mathbf{X} \le b\}$ and $P\{a \le \mathbf{X} < b\}=P\{a < \mathbf{X} < b\}$
- If F(u) is continuous at b, then $P\{a < \mathbf{X} \le b\}=P\{a < \mathbf{X} < b\}$ and $P\{a \le \mathbf{X} \le b\}=P\{a \le \mathbf{X} < b\}$
- If F(u) is continuous at both a and b, then $F(a^+) = F(a^-)$ and $F(b^+) = F(b^-)$, and all four probabilities are equal
- That is, $P\{a < \mathbf{X} \le b\} = P\{a \le \mathbf{X} \le b\} = P\{a \le \mathbf{X} < b\} = P\{a < \mathbf{X} < b\} = F(b) - F(a)$
- Although the CDF allows us to calculate probabilities and treat all random variables in the same manner, it is often cumbersome in use
- For both discrete and continuous random variables, there are descriptions that are much easier to use

Discrete Random Variables

- The CDF of a discrete random variable is a series of steps



- It suffices to know the height and location of each step
- A *discrete* random variable **X** takes on discrete values $u_1, u_2, u_3, \ldots u_n, \ldots$ where $u_1 < u_2 < u_3 < \ldots < u_n < \ldots$
- The probability mass function of **X** tells us the probabilities with which these values occur
- The probability mass function (pmf) of **X**, also known as the discrete density function of **X**, is denoted by $p(u)$ or $p_{\mathbf{X}}(u)$
- $p(\bullet)$ is a *function* with *domain* $\{u_1, u_2, u_3, \ldots\}$ and *range* $[0,1]$
- $p(u_i) = P\{\mathbf{X} = u_i\}$, $i = 1, 2, 3, \ldots$
- Properties of a pmf:
  1. $p(u_i) \geq 0$ for all i
  2. $\sum_i p(u_i) = 1$
- $F(u) = \sum_{u_i \leq u} p(u_i)$
- The pmf specifies that a *probability mass* of $p(u_i)$ is located at u_i, $i = 1, 2, 3, \ldots$
- Masses are nonnegative, and the total probability mass = 1



- $p(u_i) = P\{\mathbf{X} = u_i\}$, $i = 1, 2, 3, \ldots$ is a system of *point masses*
- $F(u)$ = total mass to the left of (and including) the point u

- Later, we shall give meaning to the concepts of center of mass and moment of inertia of these point masses
- **Example:** **X** is the number of heads when a biased coin with P(H) = p is tossed twice
- **X** takes on values 0, 1, 2 with probabilities q^2, 2pq and p^2
- The pmf of **X** is $p(0) = q^2$, $p(1) = 2pq$, $p(2) = p^2$
- $F(u) = \sum_{u_i \le u} p(u_i)$
- $p(0) = q^2$, $p(1) = 2pq$, $p(2) = p^2$ • $F(u) = \begin{cases} 0, & \text{for } u < 0 \\ q^2 & \text{for } 0 \le u < 1 \\ 1-p^2 & \text{for } 1 \le u < 2 \\ 1 & \text{for } 2 \le u \end{cases}$
- Previously encountered pmfs
- **Binomial random variable** with parameters (N, p) takes on values 0, 1, 2, … , N
- $p(k) = P\{\mathbf{X} = k\} = \binom{N}{k} p^k q^{N-k}$
- **Geometric random variable** with parameter p**:** For k = 1, 2, … $p(k) = P\{\mathbf{X} = k\} = pq^{k-1}$
- Negative binomial random variable with parameters (r, p) takes on values r, r+1, …
  For k = r, r+1, … , $p(k) = P\{\mathbf{X} = k\} = \binom{k-1}{r-1} p^r q^{k-r}$
- **Poisson random variable** with parameter λ takes on values 0, 1, 2, …
- $p(k) = \exp(-\lambda)\frac{\lambda^k}{k!}$, k = 0, 1, 2, …
- For large N and small p, the (N, p) binomial random variable is approximated by the Poisson random variable with parameter λ = Np

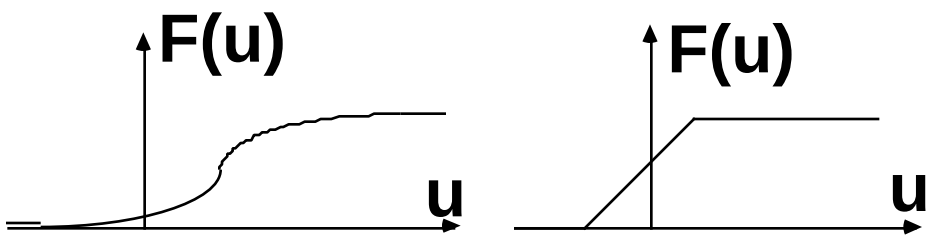
$$\sum_{k=0}^{\infty} p(k) = \sum_{k=0}^{\infty} \exp(-\lambda)\frac{\lambda^k}{k!} = \exp(-\lambda)\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \exp(-\lambda)\exp(\lambda) = 1$$

- The Poisson pmf arises in many applications
- Read pp. 154-162 of Ross and Example 8 by yourself
- A discussion of the Poisson pmf (and its relation to some continuous random variables is deferred till later in the course)

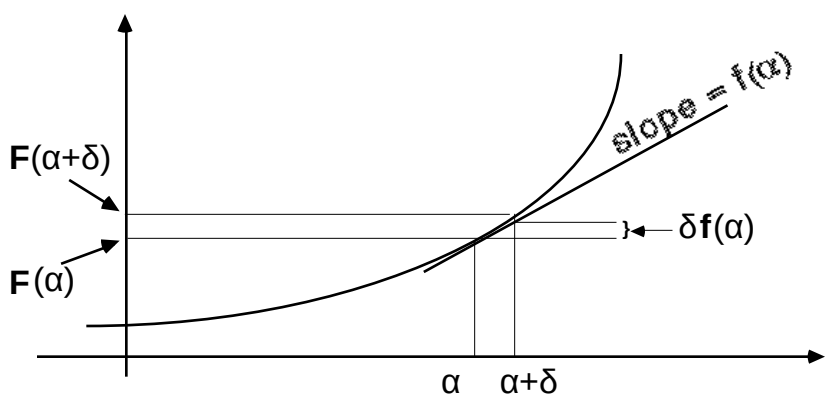
- On exams, I expect you to know what is meant by the phrases “Binomial random variable with parameters (N, p)”, “Geometric random variable with parameter p”, and “Poisson random variable with parameter λ”. You should know what their pmfs are (as well as their means and variances — concepts defined a little later in the course)

Continuous Random Variables

- Continuous random variables can take on all possible values in (or an interval, or a union of intervals in)
- The CDF F(u) of a continuous random variable is continuous everywhere
- F(u) is continuous for all u, $-\infty < u < \infty$
- A random variable is said to be continuous if its CDF F(u) is absolutely continuous, i.e. differentiable everywhere (except possibly at some finite set of points, where it is continuous but not differentiable)



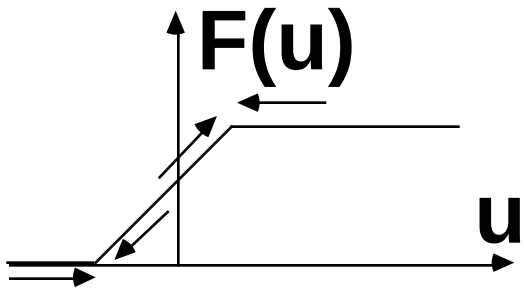
- $P\{a < \mathbf{X} \le b\} = F(b) - F(a)$
- $P\{a \le \mathbf{X} \le b\} = F(b) - F(a)$ also because F(u) is continuous everywhere
- Let $b = a + \delta$ where δ is small
- $P\{a \le \mathbf{X} \le a + \delta\} = F(a + \delta) - F(a) = ?$
- **Digression:** F(u) is said to be differentiable at the point α if $\lim_{\delta \to 0} \frac{F(\alpha + \delta) - F(\alpha)}{\delta}$ exists
- The value of the limit is called the derivative of F(u) at the point $u = \alpha$
- $f(u) = \frac{dF(u)}{du}$ or $\frac{dF}{du}$
- $f(\alpha) = \lim_{\delta \to 0} \frac{F(\alpha + \delta) - F(\alpha)}{\delta}$ = slope of F at point α
- If δ is small, then $F(\alpha + \delta) - F(\alpha) \approx f(\alpha) \times \delta$
- The approximation improves as δ gets smaller and smaller
- If δ is small, then $F(\alpha + \delta) - F(\alpha) \approx f(\alpha) \times \delta$



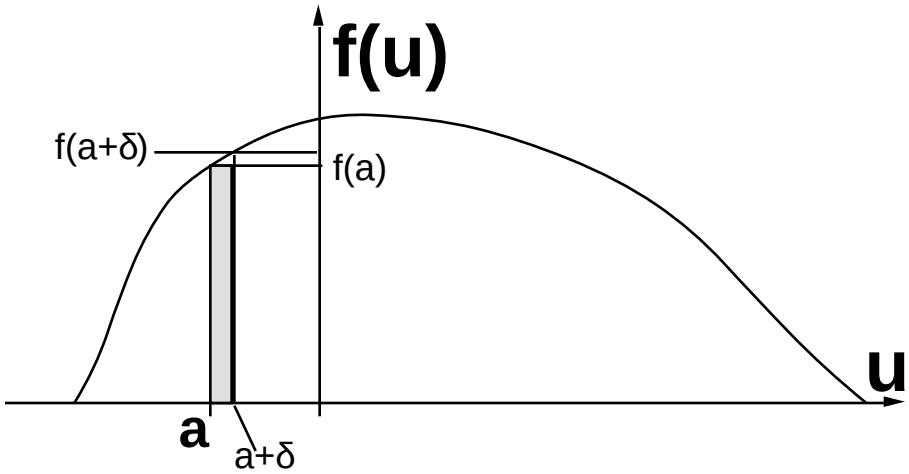
- End of digression

- If the continuous random variable **X** has CDF F(u), then $f(u) = \frac{dF(u)}{du}$ is called the *probability density function* (pdf) of **X**. Note that $f(u) \geq 0$.
- $P\{a \leq \mathbf{X} \leq a + \delta\} = F(a + \delta) - F(a) \approx f(a) \times \delta$
- The pmf of a discrete random variable describes the locations of a set of point masses
- If p(0.2) = 0.3, the random variable takes on value 0.2 with probability 0.3
- The total probability mass equals 1
- In contrast, $P\{\mathbf{X} = u\} = 0$ for *all* u for a continuous random variable **X**
- Remember F(u) is continuous everywhere so that $P\{\mathbf{X} = u\} = F(u^+) - F(u^-) = 0$ for all u, $-\infty < u < \infty$
- Where did the probability disappear?
- For continuous random variables, it makes no sense to ask for the probability of occurrence of a *specific value* for **X**
- It makes sense to ask for the probability that **X** has value in some *interval*, and these probabilities are nonzero
- For small δ, $P\{a \leq \mathbf{X} \leq a + \delta\} = F(a + \delta) - F(a) \approx f(a) \times \delta > 0$ if $f(a) > 0$
- With continuous random variables, there are no *point* masses; instead the total probability mass is spread along the real line with variable *density*
- The density of the probability mass at different points is given by the *probability density function* (pdf)
- If f(1) is large, the mass is very dense there
- If f(2) is small, the mass is not very dense there

- The value of f(u) is *not the probability mass* at the point u; it is the *density* of the mass at the point u
- The units for f(u) are mass/unit length, e.g. oz./in.
- To get a mass, you *must* "multiply" by a length, e.g. $P\{a \leq \mathbf{X} \leq a + \delta\} \approx f(a) \times \delta$
- Suppose that F(u) is not differentiable at α
- There can only be a finite number of such points
- In such cases, $f(\alpha)$ is taken to be equal to the "derivative from the right" or the "derivative from the left"



- $P\{a \le \mathbf{X} \le a + \delta\} = F(a + \delta) - F(a) \approx f(a)\times\delta$
- $P\{a \le \mathbf{X} \le a + \delta\} \approx f(a)\times\delta$ = area of rectangular strip of width δ and height f(a)

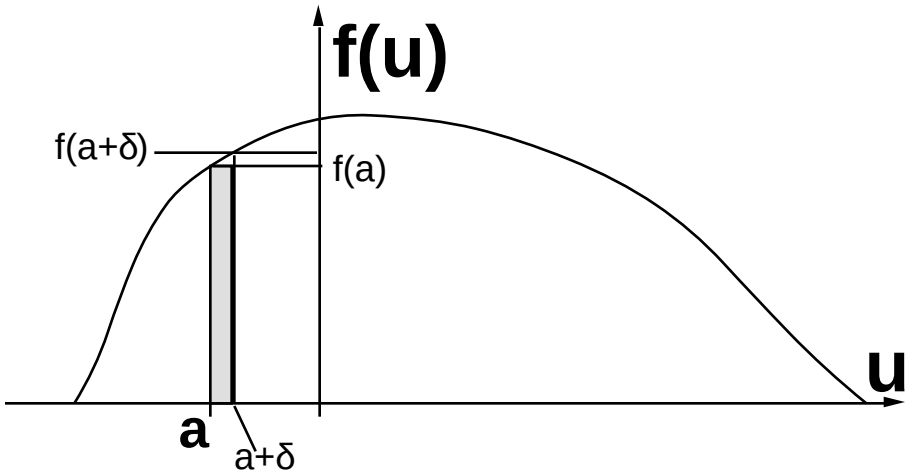


- For continuous random variables, it makes no sense to ask for the probability of occurrence of a *specific value* for **X**
- It makes sense to ask for the probability that **X** has value in some *interval*, and these probabilities are nonzero
- For small δ, $P\{a \le \mathbf{X} \le a + \delta\} = F(a + \delta) - F(a) \approx f(a)\times\delta > 0$ if $f(a) > 0$
- With continuous random variables, there are no *point* masses; instead the total probability mass is spread along the real line with variable *density*
- The density of the probability mass at different points is given by the *probability density function* (pdf)
- The value of f(u) is *not the probability mass* at the point u; it is the *density* of the mass at the point u
- The units for f(u) are mass/unit length, e.g. oz./in.
- To get a mass, you *must* "multiply" by a length, e.g. $P\{a \le \mathbf{X} \le a + \delta\} \approx f(a)\times\delta$
- $f(u) = \dfrac{dF(u)}{du}$
- So, $F(u) = \int f(u)du$??
- Not quite
- Let us first try to express $P\{a \le \mathbf{X} \le b\}$ in terms of the pdf f(u)
- $P\{a \le \mathbf{X} \le b\}$

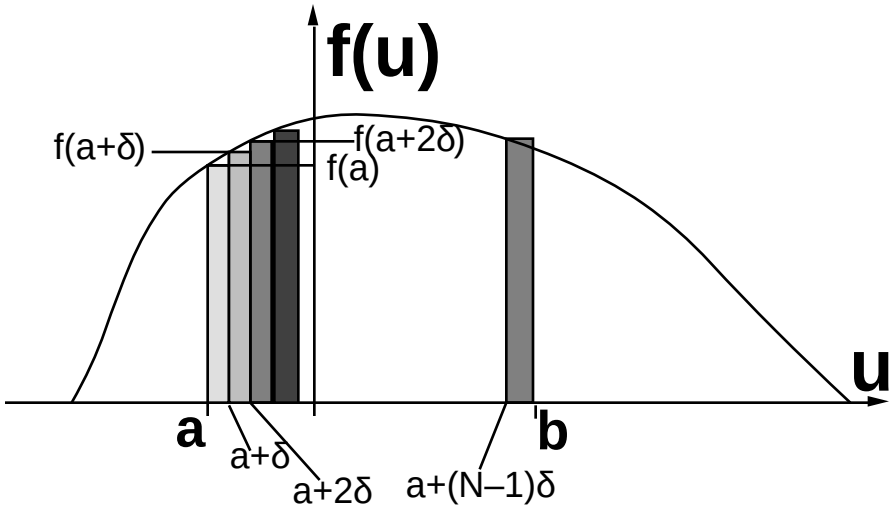
$$= P\{a \le \mathbf{X} \le a+\delta\} + P\{a+\delta \le \mathbf{X} \le a+2\delta\} + P\{a+2\delta \le \mathbf{X} \le a+3\delta\} + \ldots$$

$$+ P\{a+(N-1)\delta \le \mathbf{X} \le a+N\delta\} \approx \sum_{i=0}^{N-1} f(a+i\delta)\times\delta \text{ where } a+N\delta = b, \text{ i.e., } N = (b-a)/\delta$$

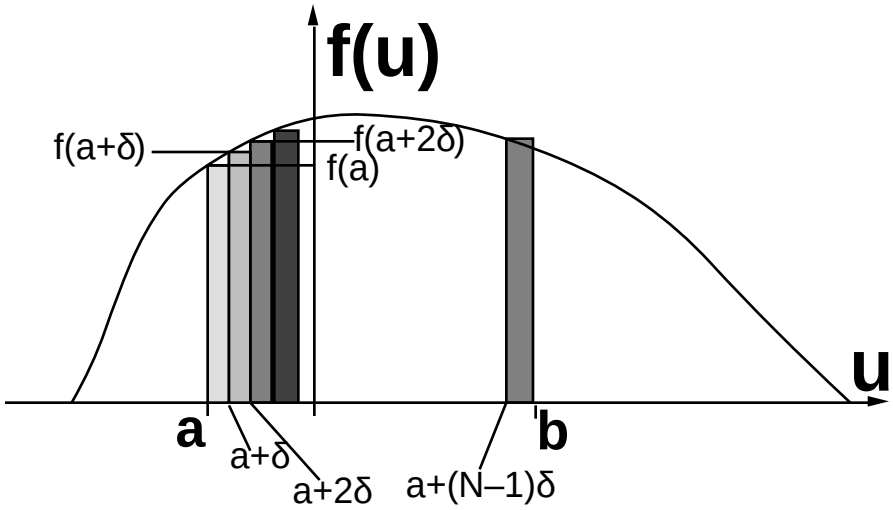
- $P\{a \le \mathbf{X} \le a + \delta\} \approx f(a)\times\delta$ = area of rectangular strip of width δ and height f(a) as shown



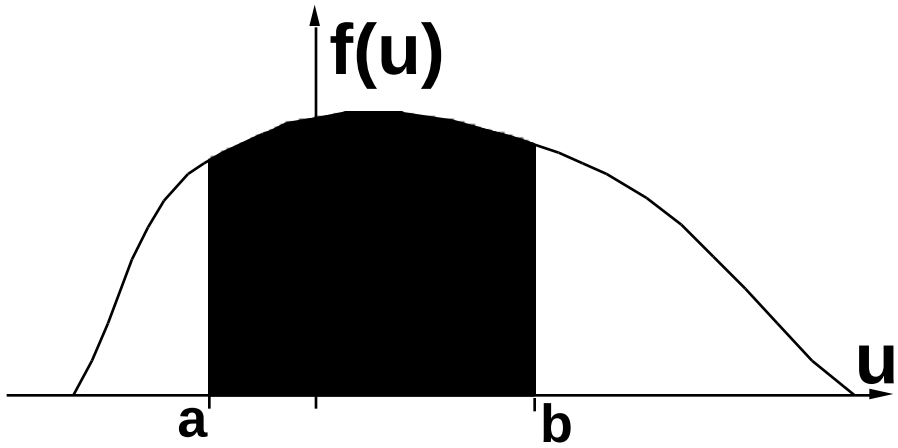
- $P\{a \le \mathbf{X} \le b\} \approx \sum_{i=0}^{N-1} f(a+i\delta)\times\delta$



- $P\{a \le \mathbf{X} \le b\}$ = limit of total area of strips as $N \to \infty$, $\delta \to 0$



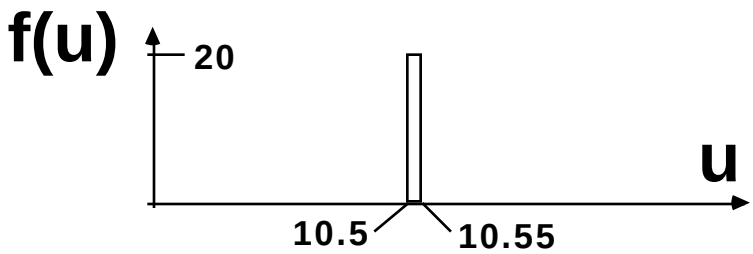
- $P\{a \le \mathbf{X} \le b\}$ = *Area* under the pdf curve from a to b



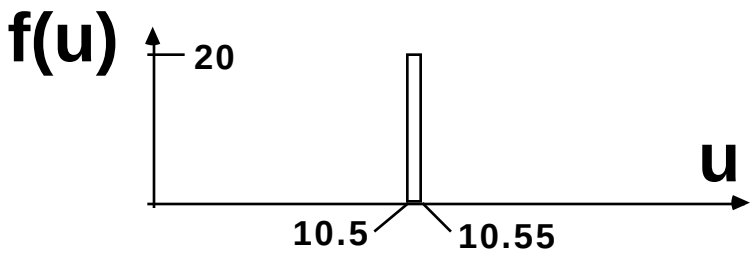
- $P\{a \le \mathbf{X} \le b\}$ = *Area* under the pdf curve from a to b
- $P\{-\infty \le \mathbf{X} \le b\} = P\{\mathbf{X} \le b\} = F(b)$ = *Area* under the pdf curve TO THE LEFT OF the point b
- $P\{-\infty \le \mathbf{X} \le \infty\} = 1$ = TOTAL AREA under the pdf curve
- **Summary:**
- f(u), the probability density function (pdf) of a continuous random variable, is the derivative of the CDF F(u)
- $f(u) = \dfrac{dF(u)}{du}, \qquad -\infty < u < \infty$
- The pdf is <u>not</u> the probability of an event
- The pdf measures how <u>dense</u> the probability mass is at any point on the real line. It is measured in units of probability mass/unit length
- To get a probability <u>mass</u>, we must multiply by a length
- For small values of δ, $P\{a \le \mathbf{X} \le a + \delta\} \approx f(a)\times\delta$
- Properties of a pdf:
- $f(u) \ge 0$
- Total AREA under the pdf curve equals 1
- Any function satisfying these two properties is a valid pdf
- How do we compute probabilities from the pdf?
- $P\{\mathbf{X} = u\} = 0$ for all u
- For small values of δ, $P\{\alpha \le \mathbf{X} \le \alpha + \delta\} \approx f(\alpha)\times\delta$

  = AREA of rectangular strip of width δ and height $f(\alpha)$
- $P\{a \le \mathbf{X} \le b\}$ = AREA under the pdf curve from a to b
- CDF F(b) = AREA under the pdf curve TO THE LEFT OF the point b
- **Example:** Is the following a valid pdf?

  $f(u) = \begin{cases} 20, & 10.5 \le u \le 10.55, \\ 0, & \text{elsewhere.} \end{cases}$
- ALWAYS BEGIN WITH A SKETCH of the pdf curve
- Things are a LOT easier when you can visualize them with a sketch in front of you

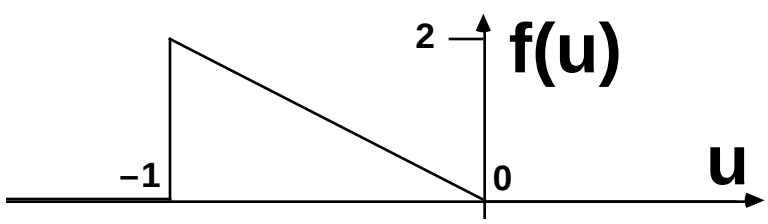


- Is f(u) ≥ 0 for all u? Yes!
- Is the total area under the pdf curve = 1? Yes! AREA = 20×(10.55–10.50) = 1
- So this is a valid pdf
- How can this be a valid pdf? Its value exceeds 1, and probabilities cannot exceed 1
- f(u) is <u>not</u> a probability; it is the <u>density</u> of the probability mass at u

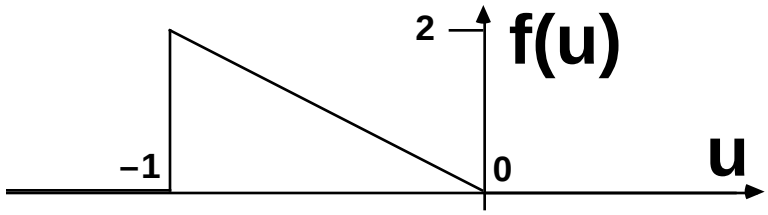


- The probability mass is *very dense* in the vicinity of 10.525
- One cannot have extensive regions of high density because the total probability mass is limited to 1
- **Example:** Is this a valid pdf?

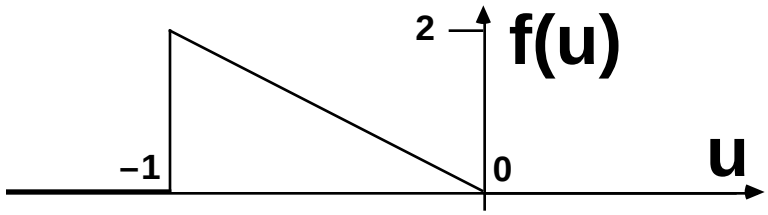
  $$f(u) = \begin{cases} -2u, & -1 \le u \le 0, \\ 0, & \text{elsewhere.} \end{cases}$$
- No, because f(u) is negative
- Bzzzt! Wrong answer!!
- This illustrates the danger of not beginning with a sketch of the pdf. **SKETCH FIRST; THINK LATER**



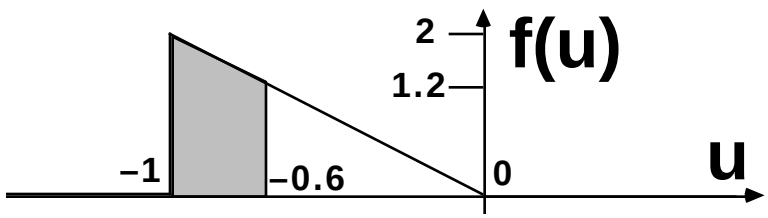
- f(u) ≥ 0
- AREA = (1/2)×base×altitude = 1
- What is the CDF of this random variable?
- We need to find F(b) = AREA under pdf curve TO THE LEFT of the point b, and we need to do it for all possible values of b
- Let's find F(b) for b = –2.1, b = –0.6, and b = 0.5 and try to generalize the answers



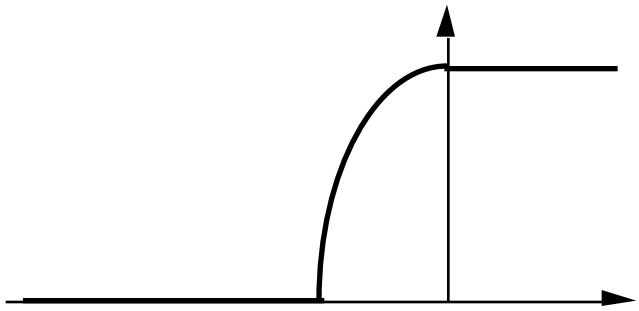
- BY INSPECTION of the graph, F(–2.1) = AREA to the left of point b = –2.1 is 0
- A little thought shows that the area is 0 no matter what b ≤ –1 we choose



- BY INSPECTION of the graph, F(0.5) = AREA to the left of point b = 0.5 is 1
- A little thought shows that the area is 1 no matter what b ≥ 0 we choose



- F(–0.6) = shaded area = 1–unshaded triangular area = $1 - (1/2)\times 0.6\times(1.2) = 1 - 0.6^2 = 0.64$
- More generally, $F(b) = 1 - b^2$ for any b, $-1 \le b \le 0$
- **Summary:** $F(u) = \begin{cases} 0, & u \le -1, \\ 1 - u^2, & -1 \le u \le 0, \\ 1, & u \ge 0. \end{cases}$
- **Quick check:** Limits? Get pdf on differentiating?

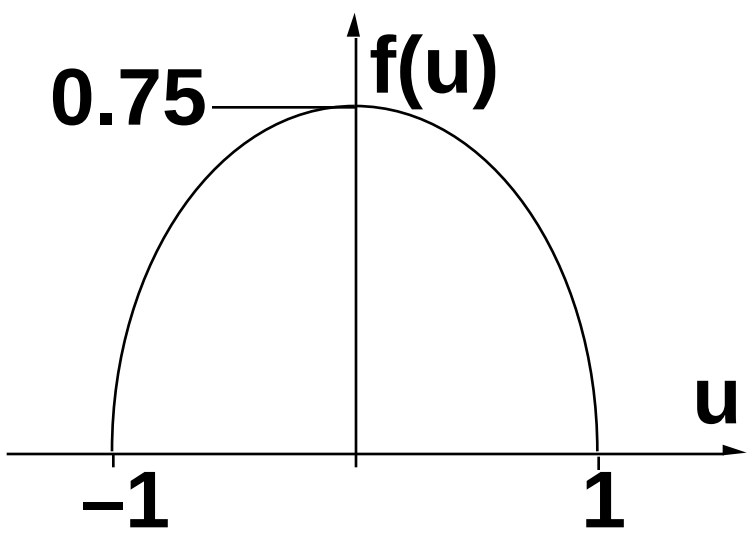


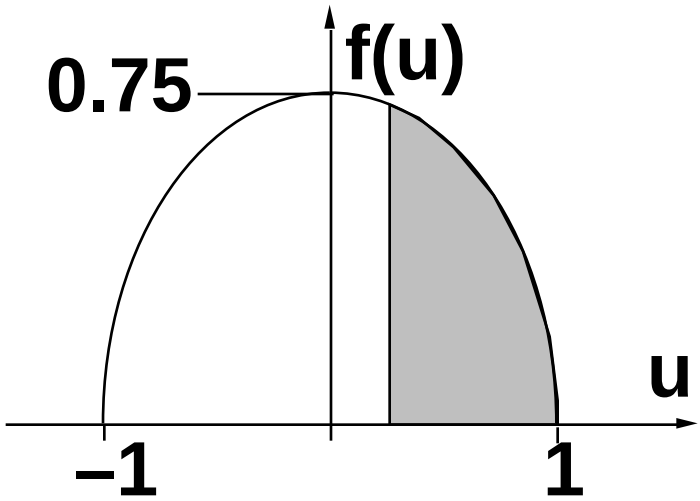
- The book says (p. 193) $F(b) = \int_{-\infty}^{b} f(u)du$
- This is **HIGHLY DANGEROUS**
- In a few weeks' time, you will be writing this as $F(u) = \int f(u)du = \int -2u\, du = -u^2$

- ALWAYS remember F(u) as the AREA under the pdf curve, NEVER as the integral of the pdf curve — you will soon confuse the integral and the antiderivative
- NAFTA =? • Remember **ANTFA: AREA, Not The Antiderivative**
- f(u), the probability density function (pdf) of a continuous random variable, is the derivative of the CDF F(u)
- $f(u) = \frac{dF(u)}{du}, \qquad -\infty < u < \infty$
- The pdf is <u>not</u> the probability of an event; we must multiply by length to get probability
- Properties of a pdf:
- $f(u) \geq 0$
- **TOTAL AREA** under the pdf curve = 1
- $P\{\mathbf{X} = u\} = 0$ for all u
- For <u>small</u> values of δ,

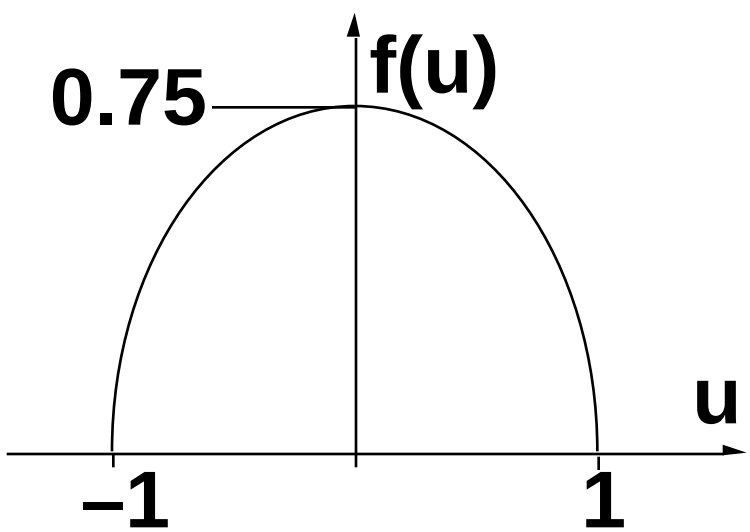
  $P\{\alpha \leq \mathbf{X} \leq \alpha + \delta\} \approx f(\alpha)\times\delta$ = **AREA** of rectangular strip of width δ and height $f(\alpha)$
- $P\{a \leq \mathbf{X} \leq b\}$ = **AREA** under the pdf curve from a to b
- CDF F(u) = **AREA** under the pdf curve TO THE LEFT OF the point u
- In finding a CDF, we must specify the value of F(u) for all u
- **Example: X** has pdf $f(u) = \begin{cases} (3/4)(1 - u^2), & |u| < 1, \\ 0, & \text{elsewhere.} \end{cases}$

  Find $P\{0.25 < \mathbf{X} < 1.25\}$.
- $P\{0.25 < \mathbf{X} < 1.25\} = \int_{0.25}^{1.25} \frac{3}{4}(1-u^2)du = \frac{3}{4}(u-u^3/3)\Big|_{0.25}^{1.25} = \frac{17}{64}$
- Bzzzt! Wrong answer
- Once again, this illustrates the danger of not beginning with a sketch of the pdf.
  SKETCH FIRST; THINK LATER





- $P\{0.25 < \mathbf{X} < 1.25\}$ = area from 0.25 to 1 under $(3/4)(1-u^2)$
- $P\{0.25 < \mathbf{X} < 1.25\} = \int_{0.25}^{1} \frac{3}{4}(1-u^2)du = \frac{3}{4}(u-u^3/3)\Big|_{0.25}^{1} = \frac{81}{256} > \frac{17}{64} = \frac{68}{256}$
- In setting up the integral correctly, we got rid of some negative area from 1 to 1.25. Hence, the net result is larger
- $P\{\mathbf{X} > 0.5 | \mathbf{X} > 0\} = \frac{P(\{\mathbf{X} > 0.5\} \cap \{\mathbf{X} > 0\})}{P\{\mathbf{X} > 0\}} = \frac{P\{\mathbf{X} > 0.5\}}{P\{\mathbf{X} > 0\}}$
- $P\{\mathbf{X} > 0.5\} = \int_{0.5}^{1} \frac{3}{4}(1-u^2)du = \frac{5}{32}$



- Note that we can *deduce* $P\{\mathbf{X} > 0\} = 1/2$ by inspection!
- $P\{\mathbf{X} > 0.5 | \mathbf{X} > 0\} = \frac{5/32}{1/2} = \frac{5}{16}$
- $P\{|\mathbf{X}|<0.5\} = 1 - 2P\{\mathbf{X}>0.5\} = \frac{11}{16}$

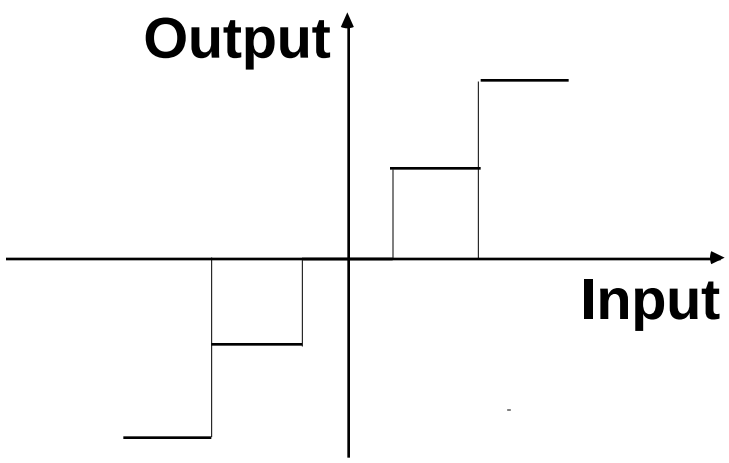
- $P\{8\mathbf{X}^2 - 6\mathbf{X} + 1 > 0\} = ?$
- $8\mathbf{X}^2 - 6\mathbf{X} + 1 = (4\mathbf{X} - 1)(2\mathbf{X} - 1)$
- $8\mathbf{X}^2 - 6\mathbf{X} + 1 > 0$ if and only if both factors are positive or both are negative
- $8\mathbf{X}^2 - 6\mathbf{X} + 1 > 0$ if $4\mathbf{X} - 1 > 0$ *AND* $2\mathbf{X} - 1 > 0$
  or if $4\mathbf{X} - 1 < 0$ *AND* $2\mathbf{X} - 1 < 0$
- $4\mathbf{X} - 1 > 0$ if $\mathbf{X} > 0.25$ while $2\mathbf{X} - 1 > 0$ if $\mathbf{X} > 0.5$
  Thus, $4\mathbf{X} - 1 > 0$ and $2\mathbf{X} - 1 > 0$ both hold if and only if $\mathbf{X} > 0.5$
- $P\{8\mathbf{X}^2 - 6\mathbf{X} + 1 > 0\} = P(\{\mathbf{X} > 0.5\} \quad \{\mathbf{X} < 0.25\}) = P\{\mathbf{X} > 0.5\} + P\{\mathbf{X} < 0.25\}$
- $P\{8\mathbf{X}^2 - 6\mathbf{X} + 1 > 0\} = P\{\mathbf{X} > 0.5\} + P\{\mathbf{X} < 0.25\} = \frac{5}{32} + (1 - P\{\mathbf{X} > 0.25\})$
  $= \frac{5}{32} + (1 - \frac{81}{256}) = \frac{215}{256}$
- This is how we use the basic result: $P\{a \le \mathbf{X} \le b\}$ = AREA under the pdf curve from a to b to find probabilities.

- A primer on surviving ECE 313
- NEVER, EVER, remember the integral formulas given by Ross on pp. 192-193
- **ALWAYS SKETCH** the pdf **BEFORE DOING ANYTHING**
- Probabilities are given by areas under the pdf curve that you have drawn. Mark the areas that you need to compute by shading or cross-hatching.
- Only NOW should you attempt to compute the area. Maybe it will be obvious from the sketch, maybe simple calculations based on high-school geometry will suffice, maybe an integral will have to be computed. But the limits on the integral should be obvious from the **SKETCH**
- **ON EXAMS, NO PARTIAL CREDIT** is given if you have not sketched the pdf first. If you believe that you are smart enough to do it all in your head, then you should be confident enough to solve the problem correctly all the way to the last period.

FUNCTIONS OF RANDOM VARIABLES

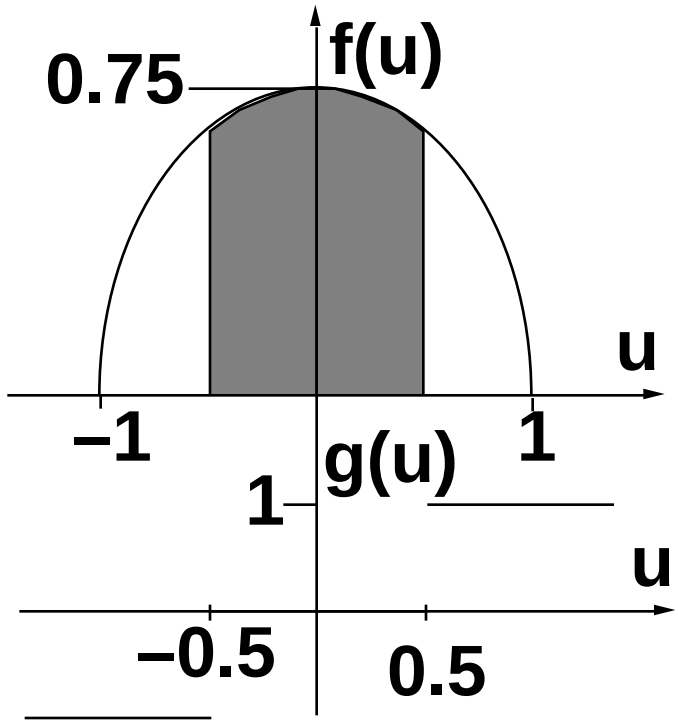
- A random variable **X** denotes the numerical outcome of a random experiment
- As the experiment is repeated over and over, we observe different values of **X**
- We don't know which value will be observed except in a probabilistic sense
- Let g(u) denote a real-valued function of a real variable u
- **Example:** $g(u) = u^2$ maps the number u onto its square, so that if u = 5, g(u) = 25
- Each time the experiment is performed, we observe some random <u>number</u> and can apply g(•) to this number
- **Example:** A coin is tossed twice and we observe the number of heads, which is either 0 or 1 or 2. We apply $g(u) = u^2 + 1$ to the number observed and get 1 or 2 or 5
- We thus have a new set of numbers that we can observe
- A random variable **X** is a function $\mathbf{X}: \Omega \rightarrow$
- Every outcome in the sample space Ω is mapped onto a real number by the function **X**
- The real number assigned to $\omega \quad \Omega$ is denoted by $\mathbf{X}(\omega)$
- Remember that there is nothing random about the function or map **X** itself. This function is fixed and nonrandom; ω is always mapped to the same number. Randomness arises from the fact that we cannot predict exactly which outcome $\omega \quad \Omega$ is going to occur
- If ω is the outcome of the experiment, we observe the number $\mathbf{X}(\omega)$
- Apply the function g(•) to this number $\mathbf{X}(\omega)$ to get a new number $g(\mathbf{X}(\omega))$
- We have constructed another function from Ω to
- New function maps $\omega \quad \Omega$ onto the real number $g(\mathbf{X}(\omega))$
- g(•) is a known fixed function
- The mapping $\mathbf{X}: \Omega \rightarrow$ is a fixed mapping
- $g(\mathbf{X}(\bullet))$ is a fixed mapping from Ω to ; it is also a random variable
- Let **Y** denote this random variable
- $\mathbf{X}: \Omega \rightarrow$ maps $\omega \quad \Omega$ onto the real number $\mathbf{X}(\omega)$
- $\mathbf{Y}: \Omega \rightarrow$ maps $\omega \quad \Omega$ onto the real number
  $\mathbf{Y}(\omega) = g(\mathbf{X}(\omega))$
- For obvious reasons, we write $\mathbf{Y} = g(\mathbf{X})$
- **Y** is said to be a function (the function g(•), of course!) of the random variable **X**
- The range of **Y** is the image under g(•) of the range of **X**
- To find the value taken on by **Y** on a particular trial of the experiment, just apply g(•) to the value taken on by **X**
- **Fundamental problem:** given g(•) and the CDF (or pmf or pdf) of **X**, find the CDF (or pmf or pdf) of $\mathbf{Y} = g(\mathbf{X})$
- Simplest case: **X** is a discrete random variable with values $u_1, u_2, u_3, \ldots, u_n, \ldots$ where $u_1 < u_2 < u_3 < \ldots < u_n < \ldots$
- Then, $\mathbf{Y} = g(\mathbf{X})$ is also a discrete random variable taking on values $v_1 = g(u_1), v_2 = g(u_2), \ldots, v_n = g(u_n), \ldots$
- **Caveat:** $v_1, v_2, \ldots, v_n, \ldots$ may not always be in increasing order like the u_i's

- **Caveat:** Some v_i's may have the same value
- **Example:** If **X** takes on values –1, 0, 1, then **Y** = $\mathbf{X}^2$ takes on values +1, 0, +1 respectively
- Finding the pmf of **Y** from the pmf of **X** is easy
- $p_{\mathbf{X}}(u_i) = P\{\mathbf{X} = u_i\}$
- $p_{\mathbf{Y}}(v_j) = \sum_{i:\, g(u_i) = v_j} p_{\mathbf{X}}(u_i)$
- Sum the probabilities of all u_i such that $g(u_i) = v_j$
- Example: If **X** takes on values –1, 0, +1, then **Y** = $\mathbf{X}^2$ takes on values +1, 0, +1 respectively
- $p_{\mathbf{Y}}(+1) = p_{\mathbf{X}}(-1) + p_{\mathbf{X}}(+1)$
  $p_{\mathbf{Y}}(0) = p_{\mathbf{X}}(0)$
- CDFs can be obtained, but are messier (as usual)
- If **X** is a continuous random variable, **Y** = g(**X**) can be a discrete random variable or a mixed random variable or a continuous random variable
- **Special Case:** **X** continuous: **Y** discrete
- **X** is the input and **Y** the output of a <u>quantizer</u>



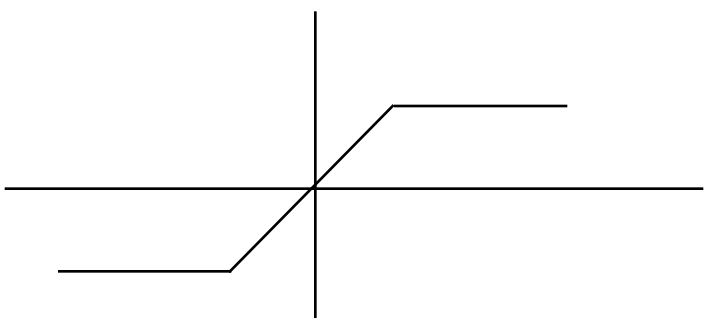
- The pmf of **Y** is easily obtained from the pdf of **X**
- If all values of **X** in the range (a,b) are quantized to **Y** = v, then $p_{\mathbf{Y}}(v) = P\{a < \mathbf{X} < b\}$
  = AREA under the pdf curve $f_{\mathbf{X}}(u)$ from a to b
- **Example:** **X** has pdf $f_{\mathbf{X}}(u) = \begin{cases} (3/4)(1 - u^2), & |u| < 1, \\ 0, & \text{elsewhere.} \end{cases}$

$$g(u) = \begin{cases} -1, & u \le -0.5, \\ 0, & |u| < 0.5, \\ +1, & u \ge +0.5. \end{cases}$$



- $p_{\mathbf{Y}}(0) = P\{|\mathbf{X}| < 0.5\} = \frac{11}{16}$. $\qquad p_{\mathbf{Y}}(+1) = P\{\mathbf{X} \geq 0.5\} = \frac{5}{32}$

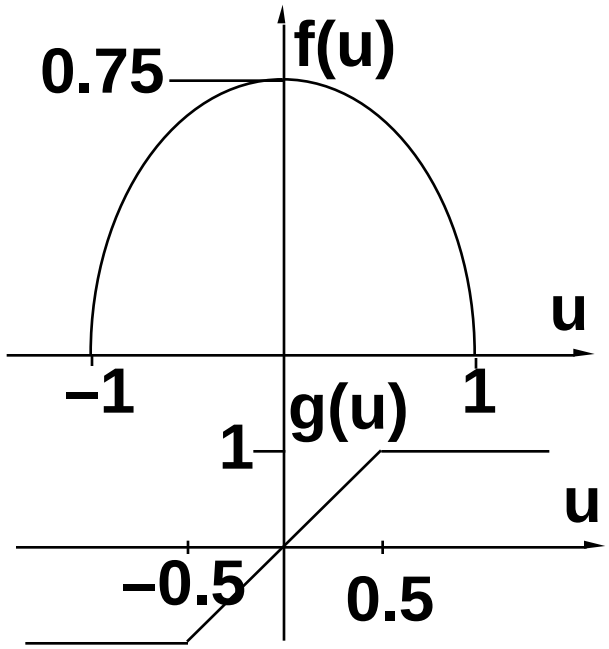
  By symmetry, $p_{\mathbf{Y}}(-1) = P\{\mathbf{X} \leq -0.5\} = \frac{5}{32}$ also!
- **Special Case:** **X** continuous: **Y** mixed
- This case arises, for example, when **X** is the input and **Y** the output of a <u>saturating amplifier</u>



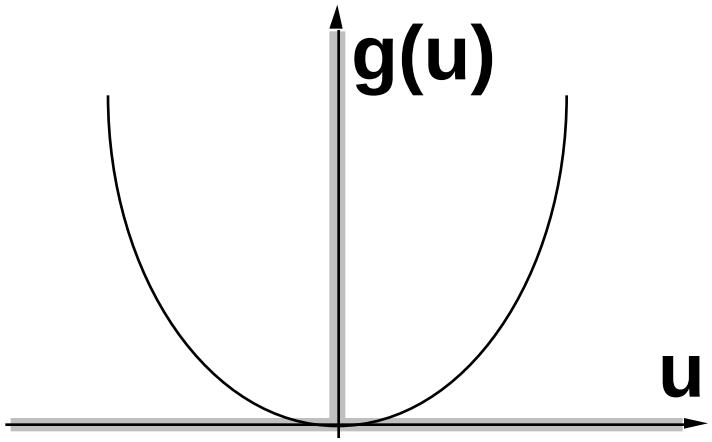
- In this case, it is best to find the CDF of the mixed random variable
- **Example:** **X** has pdf

$$f_{\mathbf{X}}(u) = \begin{cases} (3/4)(1 - u^2), & |u| < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

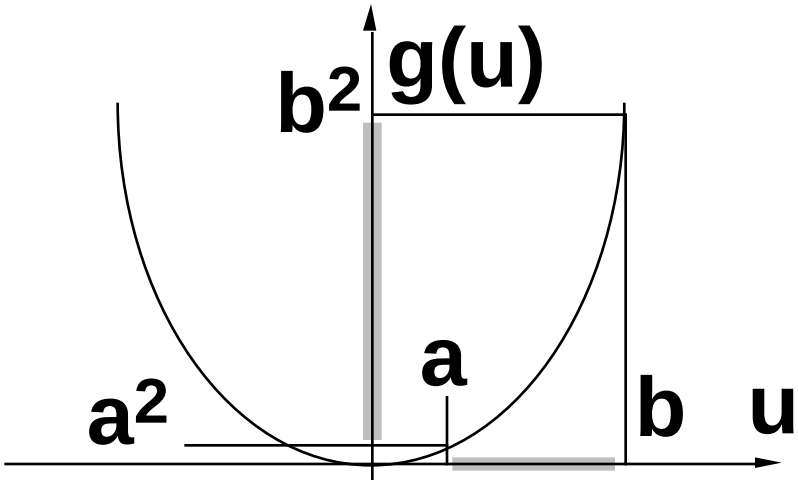
$$g(u) = \begin{cases} -1, & u \leq -0.5, \\ 2u, & |u| < 0.5, \\ +1, & u \geq +0.5. \end{cases}$$



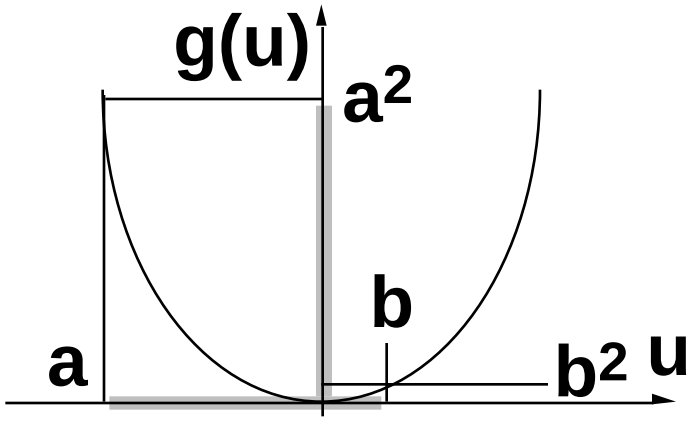
- $\mathbf{Y} = g(\mathbf{X})$ is a mixed random variable taking on values in the range [–1, +1]
- $F_{\mathbf{Y}}(v) = 0$ if $v < -1$ and also, $F_{\mathbf{Y}}(v) = 1$ if $v \geq +1$ (Why?)
  $F_{\mathbf{Y}}(-1) = P\{\mathbf{X} \leq -0.5\} = \frac{5}{32}$
  We can find $F_{\mathbf{Y}}(v)$ for v in the range $-1 < v < +1$ as follows:
- $F_{\mathbf{Y}}(0.5) = P\{\mathbf{Y} \leq 0.5\} = P\{2\mathbf{X} \leq 0.5\} = P\{\mathbf{X} \leq 0.25\} = 1 - P\{\mathbf{X} > 0.25\} = 1 - \frac{81}{256} = \frac{175}{256}$
  Generally, for any v, $-1<v<1$, $F_{\mathbf{Y}}(v) = P\{\mathbf{Y} \leq v\} = P\{2\mathbf{X} \leq v\} = P\{\mathbf{X} \leq v/2\} = F_{\mathbf{X}}(v/2)$
- $F_{\mathbf{X}}(v/2) =$ AREA under pdf curve $f_{\mathbf{X}}(u)$ upto point v/2 = $\int_{-1}^{v/2} \frac{3}{4}(1-u^2)du = \frac{1}{2} + \frac{3v}{8} - \frac{v^3}{32}$
- $\frac{1}{2} + \frac{3v}{8} - \frac{v^3}{32} = \frac{5}{32}$ as $v \to -1$
- $\frac{1}{2} + \frac{3v}{8} - \frac{v^3}{32} = \frac{27}{32}$ as $v \to +1$
- Summary: $\mathbf{Y} = g(\mathbf{X})$ is a mixed random variable taking on values in the range [–1, +1]
- $F_{\mathbf{Y}}(v) = \begin{cases} 0, & v < -1, \\ \frac{1}{2} + \frac{3v}{8} - \frac{v^3}{32}, & -1 \leq v < 1, \\ 1, & v \geq 1 \end{cases}$
- Note the jump-discontinuities at $v = -1$ and $v = +1$
- **Special Case:** $\mathbf{X}$ continuous: $\mathbf{Y}$ continuous
- How to find $f_{\mathbf{Y}}(v)$ from $f_{\mathbf{X}}(u)$?
- Use knowledge of $g(\bullet)$ and the range of $\mathbf{X}$ to find the range of $\mathbf{Y}$, that is, the minimum value α and maximum value β taken on by $\mathbf{Y}$
- **Example:** If $\mathbf{Y} = \mathbf{X}^2$ where $\mathbf{X}$ takes on values in the range $(-\infty, \infty)$, then $\mathbf{Y}$ takes on values in the range $[0, \infty)$



- **Example:** If $\mathbf{Y} = \mathbf{X}^2$ where $0 < a < \mathbf{X} < b$, then $a^2 < \mathbf{Y} < b^2$

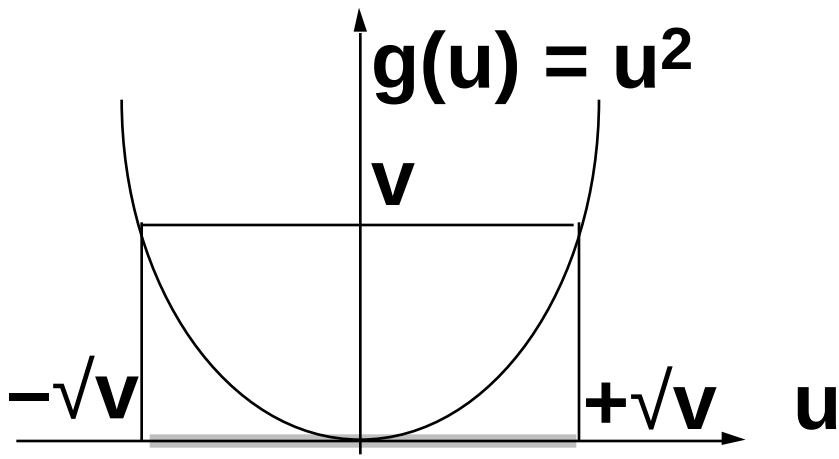


- **Example:** If $\mathbf{Y} = \mathbf{X}^2$ where
  $\mathbf{X}$ (a,b) with $a < 0 < b$, then $\mathbf{Y}$ $[0, \max\{a^2,b^2\})$



- $\mathbf{Y}$ has range (α, β) or $[\alpha,\beta)$ or...
- $F_{\mathbf{Y}}(v) = 0$ for all $v < \alpha$
- $F_{\mathbf{Y}}(v) = 1$ for all $v > \beta$
- $f_{\mathbf{Y}}(v) = 0$ for all v (α, β) or ...

- For any v in the range, find $F_{\mathbf{Y}}(v)$ in terms of $F_{\mathbf{X}}(\bullet)$
- Differentiate to get $f_{\mathbf{Y}}(v)$
- **Example:** $\mathbf{Y} = \mathbf{X}^2 \geq 0$ always
- Let v denote a number ≥ 0
- $F_{\mathbf{Y}}(v) = P\{\mathbf{Y} \leq v\} = P\{\mathbf{X}^2 \leq v\}$



- $F_{\mathbf{Y}}(v) = P\{\mathbf{Y} \leq v\} = P\{\mathbf{X}^2 \leq v\}$
  $= P\{-\sqrt{v} \leq \mathbf{X} \leq \sqrt{v}\}$
  $= F_{\mathbf{X}}(\sqrt{v}) - F_{\mathbf{X}}(-\sqrt{v})$
- $\frac{d}{dv}F_{\mathbf{Y}}(v) = f_{\mathbf{Y}}(v)$
- $\frac{d}{dv}F_{\mathbf{X}}(\sqrt{v}) = ?$ $\qquad$ $\frac{d}{dv}F_{\mathbf{X}}(-\sqrt{v}) = ?$
- We know that $\frac{d}{du}F_{\mathbf{X}}(u) = f_{\mathbf{X}}(u)$
- By the chain rule,
  $\frac{d}{dv}F_{\mathbf{Y}}(v) = \frac{d}{dv}\left[F_{\mathbf{X}}(\sqrt{v}) - F_{\mathbf{X}}(-\sqrt{v})\right] = f_{\mathbf{X}}(\sqrt{v})\times\frac{1}{2\sqrt{v}} - f_{\mathbf{X}}(-\sqrt{v})\times\frac{-1}{2\sqrt{v}} = \frac{f_{\mathbf{X}}(\sqrt{v}) + f_{\mathbf{X}}(-\sqrt{v})}{2\sqrt{v}}$
  $= f_{\mathbf{Y}}(v)$
- **Example:** If $f_{\mathbf{X}}(u) = \exp(-\pi u^2)$ for u $(-\infty,\infty)$ then $\mathbf{Y} \geq 0$
- For $v \geq 0$, $f_{\mathbf{Y}}(v) = \frac{f_{\mathbf{X}}(\sqrt{v}) + f_{\mathbf{X}}(-\sqrt{v})}{2\sqrt{v}}$

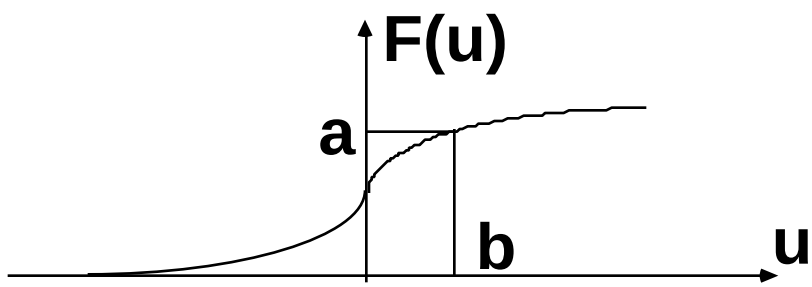
  $$f_{\mathbf{Y}}(v) = \begin{cases} v^{-1/2}\exp(-\pi v), & v \geq 0, \\ 0, & v < 0 \end{cases}$$
- **Example:** $f_{\mathbf{X}}(u) = \begin{cases} 1, & 1 < u < 2, \\ 0, & \text{elsewhere} \end{cases}$ $\qquad$ Then, $1 < \mathbf{Y} < 4$
- For any v, $1 < v < 4$, $f_{\mathbf{Y}}(v) = \frac{f_{\mathbf{X}}(\sqrt{v})+f_{\mathbf{X}}(-\sqrt{v})}{2\sqrt{v}} = \frac{f_{\mathbf{X}}(\sqrt{v})}{2\sqrt{v}}$

  $$f_{\mathbf{Y}}(v) = \begin{cases} \frac{1}{2\sqrt{v}}, & 1 < v < 4, \\ 0, & \text{elsewhere.} \end{cases}$$

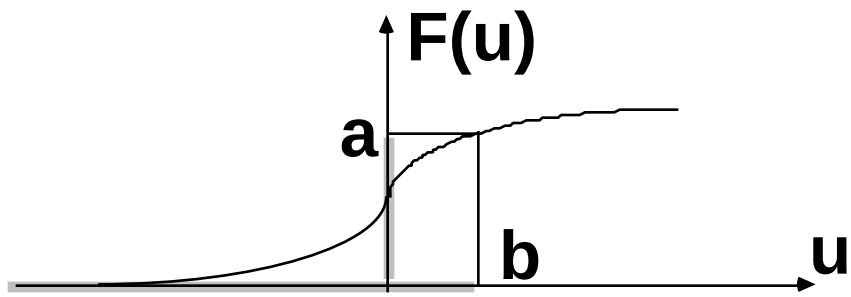
- If $f_{\mathbf{X}}(u) = \begin{cases} 0.25, & -1 < u < 3, \\ 0, & \text{elsewhere} \end{cases}$ Then, $0 \le \mathbf{Y} < 9$
- For any v, $0 \le v < 9$, $f_{\mathbf{Y}}(v) = \dfrac{f_{\mathbf{X}}(\sqrt{v}) + f_{\mathbf{X}}(-\sqrt{v})}{2\sqrt{v}}$

  If $0 \le v < 1$, then $f_{\mathbf{X}}(\sqrt{v}) = f_{\mathbf{X}}(-\sqrt{v}) = 0.25$
- If $f_{\mathbf{X}}(u) = \begin{cases} 0.25, & -1 < u < 3, \\ 0, & \text{elsewhere} \end{cases}$ Then, $0 \le \mathbf{Y} < 9$

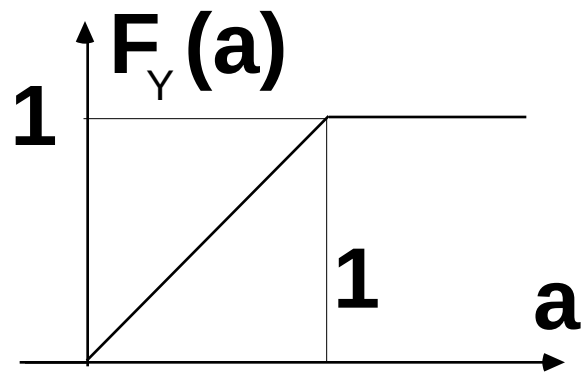
  BUT, if $v \ge 1$, then $-\sqrt{v} \le -1$, and hence $f_{\mathbf{X}}(-\sqrt{v}) = 0$
- $f_{\mathbf{Y}}(v) = \begin{cases} \dfrac{0.25}{\sqrt{v}}, & 0 \le v < 1, \\ \dfrac{0.125}{\sqrt{v}} & 1 \le v < 9, \\ 0, & \text{elsewhere.} \end{cases}$
- **Moral:** Mind your p's and q's in doing these problems
- **Guide to living a moral life**: Always begin by sketching the pdf and the function g(•)
  Figure out the range of **Y**
- For any v in the range, find $F_{\mathbf{Y}}(v)$ in terms of $F_{\mathbf{X}}(\bullet)$
- Differentiate to get $f_{\mathbf{Y}}(v)$
- **Example:** $\mathbf{Y} = g(\mathbf{X}) = a\mathbf{X} + b$ is a linear function of **X**
- $F_{\mathbf{Y}}(v) = P\{\mathbf{Y} \le v\} = P\{a\mathbf{X} + b \le v\} = \begin{cases} P\{\mathbf{X} \le (v-b)/a\}, & a > 0, \\ P\{\mathbf{X} \ge (v-b)/a\}, & a < 0, \end{cases}$
- $F_{\mathbf{Y}}(v) = \begin{cases} F_{\mathbf{X}}((v-b)/a), & a > 0, \\ 1-F_{\mathbf{X}}((v-b)/a), & a < 0, \end{cases}$ and hence $f_{\mathbf{Y}}(v) = \dfrac{1}{|a|} f_{\mathbf{X}}\left(\dfrac{v-b}{a}\right)$
- The general <u>shape</u> of the pdf of **Y** is similar to the <u>shape</u> of the pdf of **X**
- If a = 1, $f_{\mathbf{Y}}(v) = f_{\mathbf{X}}(v-b)$ is just the pdf of **X** shifted right by b
- If a = –1, $f_{\mathbf{Y}}(v) = f_{\mathbf{X}}(b-v)$ is just the the pdf of **X** "flipped over" and then shifted right by b
- If b = 0 and a > 1, $f_{\mathbf{Y}}(v) = \dfrac{1}{|a|} f_{\mathbf{X}}\left(\dfrac{v}{a}\right)$ is just the pdf of **X** s t r e t c h e d o u t along the axis, and reduced in height (Why?)
- For 0 < a < 1, the pdf is squeezed in and raised in height
- **X** is a random variable with CDF F(u)
- $F(u) = P\{\mathbf{X} \le u\}$ is a function defined for all u, $-\infty < u < \infty$
- $0 \le F(u) \le 1$ for all u
- The statement "F(b) = a" means that $P\{\mathbf{X} \le b\} = a$
- If $P\{\mathbf{X} \le b\} = a$, then F(b) = a



- F(u) is a nice well-behaved function of a real variable
- What if we take g(u) = F(u)?
- Consider the random variable $\mathbf{Y} = F(\mathbf{X})$. Then, $0 \le \mathbf{Y} \le 1$



- $\{\mathbf{Y} \le a\}$ if and only if $\{\mathbf{X} \le b\}$
- $P\{\mathbf{Y} \le a\} = P\{\mathbf{X} \le b\} = F(b) = a$
- For all a, $0 < a < 1$, $F_{\mathbf{Y}}(a) = P\{\mathbf{Y} \le a\} = a$



- $f_{\mathbf{Y}}(a) = \begin{cases} 1, & 0 < a < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- $f_{\mathbf{Y}}(a) = \begin{cases} 1, & 0 < a < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- **Y** is a <u>uniform random variable</u> on (0,1), or is <u>uniformly distributed</u> on (0,1)
- An arbitrary random variable can be changed it into a uniform random variable by the transformation $F(\mathbf{X})$
- The <u>inverse</u> transformation is very useful for <u>simulation</u>
- We wish to simulate the values of a random variable **X** with CDF F(u)
- We have available a "random number generator" whose output is a uniform random variable **Y** on (0,1)
- We know that $\mathbf{Y} = F(\mathbf{X})$
- The inverse transformation gives $\mathbf{X} = F^{-1}(\mathbf{Y})$
- To simulate **X** with specified CDF F(u), apply the function $F^{-1}(\bullet)$ to the output of a random number generator
- **Example:** We wish to simulate **X** with CDF
  $F(u) = \begin{cases} 1 - \exp(-u), & u > 0, \\ 0, & u \le 0. \end{cases}$
- If $F(\bullet)$ maps a onto b, i.e., $b = 1 - \exp(-a)$, then $\exp(-a) = 1 - b$ and hence, $a = -\ln(1 - b)$

- $\mathbf{X} = -\ln(1 - \mathbf{Y})$ where $\mathbf{Y}$ is a uniform random variable on (0,1), e.g. the output of a random number generator, has CDF and pdf given by
  $F(u) = \begin{cases} 1 - \exp(-u), & u > 0, \\ 0, & u \le 0. \end{cases}$ $f(u) = \begin{cases} \exp(-u), & u > 0, \\ 0, & u \le 0. \end{cases}$
- **Example:** We wish to simulate $\mathbf{X}$ with CDF $F(u) = \frac{1}{2} + \frac{1}{\pi}\arctan(u)$, $-\infty<u<\infty$

  and pdf $\frac{1}{\pi(1+u^2)}$, $-\infty < u < \infty$
- $F^{-1}(b) = \tan(\pi(b-0.5))$
- $\mathbf{X} = F^{-1}(\mathbf{Y}) = \tan(\pi(\mathbf{Y}-0.5))$
- $\mathbf{X} = \tan(\pi(\mathbf{Y}-0.5))$ where $\mathbf{Y}$ is a uniform random variable on (0,1), e.g. the output of a random number generator
- $\mathbf{Y}-0.5$ is uniform on $(-0.5,0.5)$
- $\pi(\mathbf{Y}-0.5)$ is uniform on $(\frac{-\pi}{2}, \frac{\pi}{2})$
- Simulation of discrete random variables based on random number generator
- $\mathbf{X} = u_1$ if $0 < \mathbf{Y} < p_1$
- $\mathbf{X} = u_2$ if $p_1 < \mathbf{Y} < p_1 + p_2$
- $\mathbf{X} = u_3$ if $p_1+p_2 < \mathbf{Y} < p_1+p_2+p_3$
  etc

Expectation: Mean and Variance

- A random variable $\mathbf{X}$ denotes the numerical outcome of a random experiment
- As the experiment is repeated over and over, we observe different values of $\mathbf{X}$
- We cannot be certain which value will be observed on any particular trial
- What is the *average* of all the values of $\mathbf{X}$ that we observe?
- Suppose $\mathbf{X}$ had value x_i on the i-th trial, $1 \le i \le N$
- Average of the observations $= \dfrac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$
- On *another* set of N trials, $\mathbf{X}$ might have values y_i, $1 \le i \le N$ and *their* average $= \dfrac{y_1 + y_2 + \ldots + y_N}{N}$ *might not be exactly the same* as $\dfrac{x_1 + x_2 + \ldots + x_N}{N}$
- For large N, such averages are nearly equal, though not necessarily identical, in value
- For large numbers of trials, we can be pretty certain about what happened in an average sense even though we cannot be sure about what happened on specific trials
- Suppose that $\mathbf{X}$ is a discrete random variable taking on values $u_1, u_2, \ldots, u_n$ with probabilities $p_1, p_2, \cdots, p_n$
- For k = 1, 2, … n, the event $\{\mathbf{X} = u_k\}$ occurs *roughly* Np_k times on N trials (N large)
- Remember that $\mathbf{X}$ had value x_i on the i-th trial, $1 \le i \le N$
- *Approximately* Np_1 of the x_i's have value u_1, Np_2 of the x_i's have value u_2, Np_3 of the x_i's have value u_3, etc. (N large)
- Average of the observations $= \dfrac{x_1 + x_2 + x_3 + \ldots + x_N}{N} \approx \dfrac{Np_1u_1 + Np_2u_2 + \ldots + Np_nu_n}{N} = p_1u_1 + p_2u_2 + \ldots + p_nu_n$
- Average is *approximately* this value (for large N)
- The *mean*, or the *average*,or the *average value*, or the *expected value*, or the *expectation* of a random variable $\mathbf{X}$ is denoted by $E[\mathbf{X}]$. Other notation often used: $E\mathbf{X}$ or μ or $\mu_{\mathbf{X}}$ or $\overline{\mathbf{X}}$
- If $\mathbf{X}$ is a discrete random variable with probability mass function (pmf) $p_{\mathbf{X}}(u)$, that is,

  $p_{\mathbf{X}}(u_i) = P\{\mathbf{X} = u_i\}$, then $E[\mathbf{X}] = \sum_i p_i u_i = \sum_i u_i p_{\mathbf{X}}(u_i)$
- Some well-known expectations
- A binomial random variable with parameters (n,p) has mean np
- A geometric random variable with parameter p has mean $\dfrac{1}{p}$
- A Poisson random variable with parameter λ has mean λ
- If $\mathbf{X}$ is a continuous random variable with pdf $f_{\mathbf{X}}(u)$, then $E[\mathbf{X}] = \int_{-\infty}^{\infty} u f_{\mathbf{X}}(u)du$
- This is just the continuous version of the sum $\sum_i u_i p_{\mathbf{X}}(u_i)$
- The event $\{a \le \mathbf{X} \le a + \delta\}$ has probability $\approx f_{\mathbf{X}}(a) \times \delta$
- This event occurs $\approx Nf_{\mathbf{X}}(a) \times \delta$ times on N trials

- Event contributes $\approx aNf_{\mathbf{X}}(a)\times\delta$ to the "sum in the numerator"
- There are multiple contributions to sum of the form $(a+i\delta)Nf_{\mathbf{X}}(a+i\delta)\times\delta$
- $E[\mathbf{X}] \approx N^{-1}\sum_{i}(a+i\delta)Nf_{\mathbf{X}}(a+i\delta)\times\delta$
- Limit of sum is the integral $E[\mathbf{X}] = \int_{-\infty}^{\infty} uf_{\mathbf{X}}(u)du$
- Uniform random variable has pdf $f_{\mathbf{X}}(u) = \begin{cases} \frac{1}{b-a}, & a < u < b \\ 0, & \text{elsewhere} \end{cases}$
- $E[\mathbf{X}] = \int_{-\infty}^{\infty} uf_{\mathbf{X}}(u)du = \int_{a}^{b} \frac{u}{b-a}du = \frac{1}{2}\times\frac{u^2}{b-a}\Big|_{a}^{b} = \frac{1}{2}\times\frac{b^2-a^2}{b-a} = \frac{b+a}{2}$
- Exponential random variable with parameter λ has pdf
  $f_{\mathbf{X}}(u) = \begin{cases} \lambda\exp(-\lambda u), & u \ge 0, \\ 0, & u < 0. \end{cases}$
- $E[\mathbf{X}] = \int_{-\infty}^{\infty} uf_{\mathbf{X}}(u)du = \int_{0}^{\infty} u\lambda\exp(-\lambda u)du = \frac{1}{\lambda}$
- $E[\mathbf{X}] = \int_{-\infty}^{\infty} uf_{\mathbf{X}}(u)du$ or $\sum_{i} u_i p_{\mathbf{X}}(u_i) = \int_{0}^{\infty} uf_{\mathbf{X}}(u)du + \int_{-\infty}^{0} uf_{\mathbf{X}}(u)du = A + B$ where $A \ge 0$ and $B \le 0$
  At least one of A and B must be finite
- $A = \int_{0}^{\infty} uf_{\mathbf{X}}(u)du$ • $B = \int_{-\infty}^{0} uf_{\mathbf{X}}(u)du$
- Both A, B finite $E[\mathbf{X}]$ finite
- $A = \infty$, B finite $E[\mathbf{X}] = \infty$
- A finite, $B = -\infty$ $E[\mathbf{X}] = -\infty$
- If $A = \infty$ and $B = -\infty$, $E[\mathbf{X}]$ does not exist, or is undefined
- **Example:** A Cauchy random variable has pdf $f(u) = \frac{1}{\pi}\frac{1}{1+u^2}$, $-\infty < u < \infty$

  $A = \int_{0}^{\infty} \frac{u}{1+u^2}du = \frac{1}{2}\ln(1+u^2)\Big|_{0}^{\infty} = \infty$ $B = \int_{-\infty}^{0} \frac{u}{1+u^2}du = -\int_{0}^{\infty} \frac{v}{1+v^2}dv = -\infty$
- Expectation = center of mass
- A point mass m at distance r from the origin has "moment about the origin" = $m\times r$
- Total moment for set of point masses m_i at distances r_i from the origin is $\sum_{i} m_i r_i$

- Center of mass is at r where $r \times \sum_i m_i = \sum_i m_i r_i$
- Total mass = $\sum_i m_i = 1$ implies that $r = \sum_i m_i r_i$
- A discrete random variable **X** has pmf consisting of probability masses p_i at u_i

  $E[\mathbf{X}] = \sum_i p_i u_i$ = location of center of mass since $\sum_i p_i = 1$
- For a continuous random variable, the probability mass $P\{a \le \mathbf{X} \le a + \delta\} \approx f_{\mathbf{X}}(a) \times \delta$ has moment $\approx a f_{\mathbf{X}}(a) \times \delta$
- In the limit, we get the integral $E[\mathbf{X}] = \int_{-\infty}^{\infty} u f_{\mathbf{X}}(u)du$ = location of center of mass
- A body is perfectly balanced about its center of mass
- There must be probability mass both to the left and to the right of E[**X**]
- If a random variable takes on values in the range (a,b) only, then $a < E[\mathbf{X}] < b$
- If the pmf or pdf is symmetric about a point, then that point is usually the center of mass, i.e. E[**X**] = point of symmetry
- However, the mass cannot be too spread out; else the moment is of the form $\infty - \infty$ and E[**X**] is undefined. An example of this is the Cauchy density which is symmetric about u = 0, but its expectation cannot be defined because the moment about the origin is of the form $\infty - \infty$.
- The expectation of a random variable is a number that summarizes the location of the probability mass on the real line
- Although E[**X**] is the average value of **X**, there may not be any $\omega \quad \Omega$ that is mapped onto E[**X**], that is, it may well be that $\mathbf{X}(\omega) \ne E[\mathbf{X}]$ for any $\omega \quad \Omega$
- E[**X**] need not be one of the possible values of **X**
- There need not be any mass at, or in the vicinity of the center of mass
- Where is the center of mass of a doughnut?
- **Example:** If **X** takes on values 0 and 1 with equal probability, its mean is 1/2, but $\mathbf{X} \ne 1/2$ on any trial
- **Expectation as fair price**
- Suppose that you win \$**X** each time you play a game of chance
- Your average winnings per game are E[**X**]
- E[**X**] is a fair price to pay to play this game
- **Example:** Let \$**X** be the amount you win in a lottery. $P\{\mathbf{X} = 1{,}000{,}000\} = 10^{-6}$ $P\{\mathbf{X} = 0\} = 1 - 10^{-6}$. Then, E[**X**] = 1 and a fair price for a lottery ticket is \$1
- More than fair price is charged to cover overhead
- E[**X**] summarizes the location of the probability mass on the real line
- Other measures of location include the *median* and the *mode*
- **Median** = number u such that F(u) = 1/2
- If F(u) = 1/2 for all u (a,b), median = (a+b)/2

- If $F(u) \neq 1/2$ for any u, then for some α, $F(\alpha^-) < 1/2$ and $F(\alpha^+) = F(\alpha) > 1/2$; median = α
- At times, the median is also called the average
- Medians are used when there are many "outliers" whose value affects the mean significantly
- "Average" salaries, income, wages etc are usually the median rather than the mean
- Median of Cauchy random variable is 0; its mean is undefined
- **Mode** of a random variable is the location of the maximum of the pmf or pdf
- Maximum may not be unique; hence mode may not be uniquely defined
- Mode is most likely value of the random variable
- The average American has had two years of college: here average means mode?
- The average American earns $17,000 per year: here average means median?
- The average American is 5' 10" tall: here average means expected value?
- $E[\mathbf{Y}] = \sum_j v_j p_{\mathbf{Y}}(v_j)$ or $\int_{-\infty}^{\infty} v f_{\mathbf{Y}}(v)dv$
- If $\mathbf{Y} = g(\mathbf{X})$, then we can calculate $p_{\mathbf{Y}}$ from $p_{\mathbf{X}}$ or $f_{\mathbf{X}}$ (whichever is appropriate) or $f_{\mathbf{Y}}$ from $f_{\mathbf{X}}$ and then use these formulas to find $E[\mathbf{Y}]$
- Computing the pmf or pdf of $\mathbf{Y}$ first and then finding $E[\mathbf{Y}]$ is too much work (and prone to calculation errors too!)
- It is far, far, better to use a magical result known as the *Law of the Unconscious Statistician* which has the wonderful acronym LOTUS
- LOTUS: If $\mathbf{Y} = g(\mathbf{X})$, then $E[\mathbf{Y}]$ can be computed directly from the pmf or pdf of $\mathbf{X}$:

  $$E[\mathbf{Y}] = \sum_j v_j p_{\mathbf{Y}}(v_j) = \sum_i g(u_i) p_{\mathbf{X}}(u_i) \qquad \text{or} \qquad E[\mathbf{Y}] = \int_{-\infty}^{\infty} v f_{\mathbf{Y}}(v)dv = \int_{-\infty}^{\infty} g(u) f_{\mathbf{X}}(u)du$$
- LOTUS says that if $\mathbf{Y} = g(\mathbf{X})$, then we can compute $E[\mathbf{Y}]$ *without finding* $p_{\mathbf{Y}}$ or $f_{\mathbf{Y}}$ first
- LOTUS *guarantees* that finding $E[\mathbf{Y}]$ as $\sum_i g(u_i) p_{\mathbf{X}}(u_i)$ yields the same result as first finding $p_{\mathbf{Y}}$ and then finding $E[\mathbf{Y}]$ via $\sum_j v_j p_{\mathbf{Y}}(v_j)$
- Formal proof of LOTUS is given in the textbook
- **Heuristic justification:** If $\mathbf{X}$ takes on values $x_1, x_2, x_3 \ldots$as we repeat the experiment, then $\mathbf{Y} = g(\mathbf{X})$ takes on values $g(x_1), g(x_2), g(x_3), \ldots$ as we repeat the experiment
- $\mathbf{Y} = g(\mathbf{X})$ takes on values $g(x_1), g(x_2), g(x_3), \ldots$ as we repeat the experiment
- The *average* value of $\mathbf{Y} = g(\mathbf{X})$ is $\dfrac{g(x_1) + g(x_2) + \ldots + g(x_N)}{N}$
- Consider a large number of trials
- $P\{a \leq \mathbf{X} \leq a + \delta\} \approx f_{\mathbf{X}}(a) \times \delta$
- Approximately $Nf_{\mathbf{X}}(a) \times \delta$ terms in numerator are $\approx g(a)$
- Event contributes $\approx g(a) N f_{\mathbf{X}}(a) \delta$ to the numerator

- Multiple contributions to sum of the form $g(a+i\delta)Nf_{\mathbf{X}}(a+i\delta)\delta$

  $E[\mathbf{Y}] \approx N^{-1}\sum_{i} g(a+i\delta)Nf_{\mathbf{X}}(a+i\delta)\times\delta$

- Limit = $E[\mathbf{Y}] = \int_{-\infty}^{\infty} g(u)f_{\mathbf{X}}(u)du$
- Proof for discrete case is similar
- **LOTUS:** If $\mathbf{Y} = g(\mathbf{X})$, then $E[\mathbf{Y}]$ can be computed *without first finding* $p_{\mathbf{Y}}$ or $f_{\mathbf{Y}}$:

  $E[\mathbf{Y}] = \sum_{i} g(u_i)p_{\mathbf{X}}(u_i)$ or $E[\mathbf{Y}] = \int_{-\infty}^{\infty} g(u)f_{\mathbf{X}}(u)du$

- **Applications of LOTUS:**

  If $\mathbf{Y} = g(\mathbf{X}) = \alpha$, (a constant), $E[\mathbf{Y}] = \sum_{i} g(u_i)p_{\mathbf{X}}(u_i) = \sum_{i} \alpha p_{\mathbf{X}}(u_i) = \alpha$

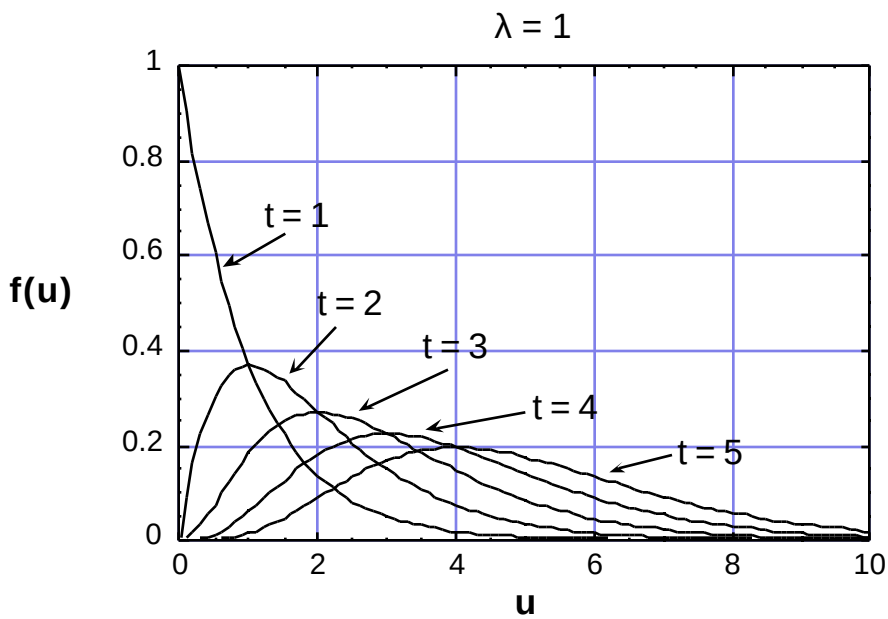
- $E[\alpha] = \int_{-\infty}^{\infty} \alpha f_{\mathbf{X}}(u)du = \alpha \int_{-\infty}^{\infty} f_{\mathbf{X}}(u)du = \alpha$
- Expectation (i.e. expected value or average value) of a constant is the constant itself
- Expectation is a *linear* operation: the expectation of a sum is the sum of the expectations
- If $\mathbf{Y} = g(\mathbf{X}) = \alpha g_1(\mathbf{X}) + \beta g_2(\mathbf{X})$, then

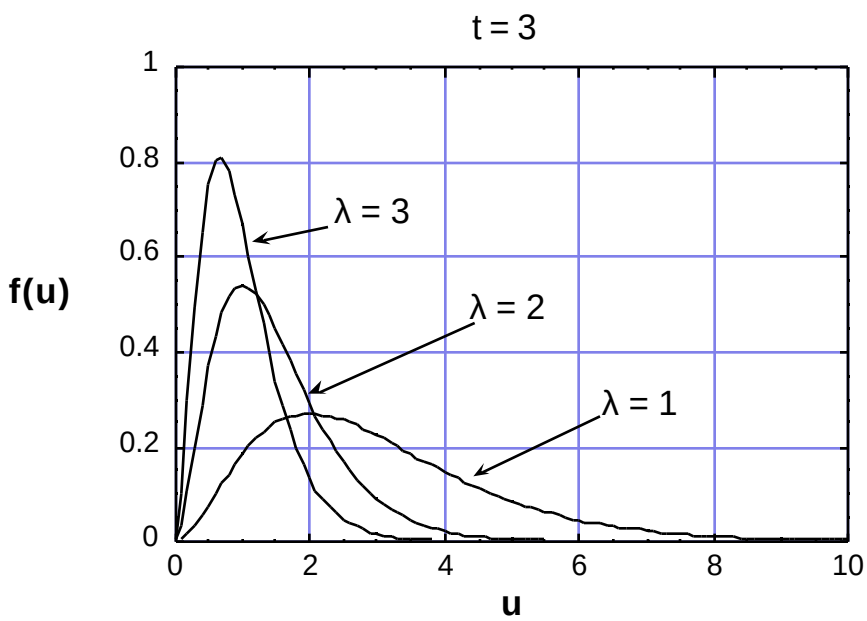
  $E[\mathbf{Y}] = \int_{-\infty}^{\infty} g(u)f_{\mathbf{X}}(u)du = \int_{-\infty}^{\infty} \{\alpha g_1(u) + \beta g_2(u)\}f_{\mathbf{X}}(u)du = \alpha E[g_1(\mathbf{X})] + \beta E[g_2(\mathbf{X})]$

- $E[\alpha\mathbf{X}] = \alpha E[\mathbf{X}]$ • $E[\mathbf{X} + \alpha] = E[\mathbf{X}] + \alpha$
- If $E[\mathbf{X}] = \mu$, then $E[\mathbf{X} - \mu] = 0$
- $E[\mathbf{X} - \mu] = \sum_{i} (u_i-\mu)p_{\mathbf{X}}(u_i)$ = moment about mean
- $E[\sum_{i} a_i\mathbf{X}^i] = \sum_{i} a_iE[\mathbf{X}^i]$
- The *variance* of a random variable $\mathbf{X}$ is $\sigma^2 = E[(\mathbf{X} - \mu)^2]$ where $\mu = E[\mathbf{X}]$
- $\sigma^2 = \sum_{i} (u_i-\mu)^2p_{\mathbf{X}}(u_i)$ or $\int_{-\infty}^{\infty} (u-\mu)^2f_{\mathbf{X}}(u)du \geq 0$
- If $\sigma^2 = 0$, all the probability mass is at μ, the center of mass, and we have a degenerate random variable (a.k.a. a constant) that always takes on value μ only!
- σ^2 is the moment of inertia about the center of mass μ
- σ = standard deviation of $\mathbf{X}$
- The variance of $\mathbf{X}$ is a measure of the "spread" of the probability mass about the mean

- The larger the variance, the greater the spread
- $\sigma^2 = E[(\mathbf{X}-\mu)^2] = E[\mathbf{X}^2 - 2\mathbf{X}\mu + \mu^2] = E[\mathbf{X}^2] - E[2\mathbf{X}\mu] + \mu^2 = E[\mathbf{X}^2] - 2\mu E[\mathbf{X}] + \mu^2$
  $= E[\mathbf{X}^2] - \mu^2$
- $\sigma^2 = E[\mathbf{X}(\mathbf{X}-1)] + \mu - \mu^2$
- These are convenient tricks for calculating σ^2
- **Example:** $\mathbf{X}$ uniform on (a,b). Using $\sigma^2 = \int_a^b (u-(a+b)/2)^2 (b-a)^{-1} du$ is messy
- $E[\mathbf{X}^2] = \int_a^b u^2 (b-a)^{-1} du = \frac{b^3-a^3}{3(b-a)}$ giving $\sigma^2 = E[\mathbf{X}^2] - \mu^2 = \frac{(b-a)^2}{12}$
- **Example:** $\mathbf{X}$ is Poisson with parameter λ. $E[\mathbf{X}] = \lambda$
- $E[\mathbf{X}(\mathbf{X}-1)] = \sum_{k=0}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2$
- $\sigma^2 = E[\mathbf{X}(\mathbf{X}-1)] + \lambda - \lambda^2 = \lambda$
- **(μ, σ^2) of the rich and famous**
- A binomial random variable with parameters (n,p) has mean np and variance np(1–p) = npq
- A geometric random variable with parameter p has mean $\frac{1}{p}$ and variance $\frac{1-p}{p^2} = \frac{q}{p^2}$
- A Poisson random variable with parameter λ has mean λ and variance λ
- A random variable uniformly distributed on (a,b) has mean $\frac{a+b}{2}$ = midpoint of interval, and variance $\frac{(b-a)^2}{12} = \frac{\text{length}^2}{12}$
- **Exponential and gamma random variables**
- $\mathbf{X}$ is said to be an exponential random variable with parameter λ if its pdf is
  $$f(u) = \begin{cases} \lambda\exp(-\lambda u), & u \ge 0, \\ 0, & u < 0. \end{cases}$$
- $\mathbf{X}$ has mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$
- $\mathbf{X}$ is a gamma random variable with parameters (t, λ); $t > 0$, $\lambda > 0$, if its pdf is
- $$f(u) = \begin{cases} \frac{\lambda\exp(-\lambda u)(\lambda u)^{t-1}}{\Gamma(t)} & u \ge 0, \\ 0, & u < 0. \end{cases}$$
- $\Gamma(t)$ = gamma function = $\int_0^{\infty} x^{t-1}\exp(-x)dx = (t-1)\Gamma(t-1) = (t-1)(t-2)\Gamma(t-2)$ and so on
- $\Gamma(1) = 1$ • $\Gamma(n) = (n-1)!$ • $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

- t = order parameter, λ = scale parameter
- Mean = $\frac{t}{\lambda}$
- Variance = $\frac{t}{\lambda^2}$

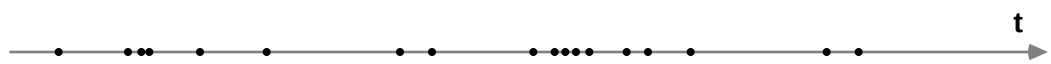




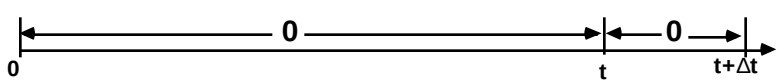
- Gamma random variable with parameters (1, λ) is an exponential random variable with parameter λ
- Gamma random variables arise in the study of "waiting times" for occurrences of random events

The Poisson random process

- Consider the occurrences of "random points" in time
- telephone off-hook signals
- jobs arriving at processor
- packets arriving at a router
- alpha particle emissions
- cars passing a checkpoint
- Each "point" occurs at some specific time instant

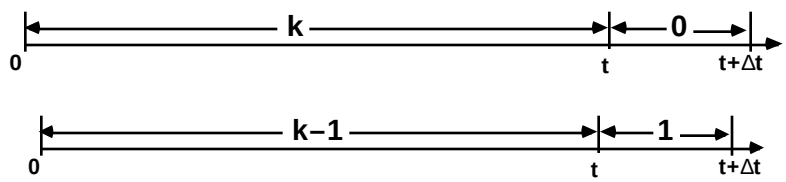


- Over a long time interval, we observe an "arrival rate" of μ points per second
- Average time interval between successive points is approximately $1/\mu$ seconds
- At each time instant t, *at most* one point can occur
- For *small* Δt, P{one point in $(t,t+\Delta t]$} = $\mu\Delta t$ and P{no point in $(t,t+\Delta t]$} = $1-\mu\Delta t$
- These are *approximations* that improve as $\Delta t \to 0$; they are nonsense for $\Delta t > 1/\mu$
- If $t_1 < t_2 \le t_3 < t_4$, then events A = {m points in $(t_1,t_2]$} and B = {n points in $(t_3,t_4]$} are independent events
- Occurrences of points in <u>disjoint</u> (i.e. non-overlapping) intervals are independent events; memoryless property
- We observe process for $t > 0$
- $P_0(t)$ = P{no points have occurred upto time t} = P{no points in $(0,t]$}
- $P_0(0)$ = P{no points in $(0,0]$} = P{no points in ø} = 1 (not 0)
- Set up and solve differential equation for $P_0(t)$
- $P_0(t+\Delta t)$ = P{no points have occurred upto time $t+\Delta t$} = $P(A\cap B)$ where
  A = {no points upto time t}; B = {no points during $(t,t+\Delta t]$}



- $P_0(t+\Delta t) = P(A\cap B) = P(A)P(B)$ since the intervals $(0,t]$ and $(t,t+\Delta t]$ are disjoint
  $= P_0(t)(1-\mu\Delta t)$
- $\dfrac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = -\mu P_0(t)$ $\qquad \dfrac{dP_0(t)}{dt} = -\mu P_0(t);$ $\qquad P_0(0) = 1$
- Solution to this differential equation is $P_0(t) = \exp(-\mu t)$
- The probability that no points occur at all during the interval $(0,t]$ should get smaller and smaller as t increases
- $P_0(\Delta t)$ = P{no point in small interval $(0,\Delta t]$} = $\exp(-\mu\Delta t) = 1 - \mu\Delta t + (\mu\Delta t)^2/2! - \ldots$
  $\approx 1 - \mu\Delta t$ for small Δt
- This too is consistent with our assumptions
- $\mathbf{X}_1$ = time of occurrence of the first point (after t = 0)
- Event $\{\mathbf{X}_1 > u\}$ occurs iff no points occur during $(0,u]$

- For $u \geq 0$, $P\{\mathbf{X}_1 > u\} = 1 - F_{\mathbf{X}_1}(u) = P_0(u) = \exp(-\mu u)$
- $F_{\mathbf{X}_1}(u) = 1 - \exp(-\mu u)$ for $u \geq 0$
- $\mathbf{X}_1$ = time of occurrence of the first random point
- $\mathbf{X}_1$ is an exponential random variable with parameter μ
- $E[\mathbf{X}_1]$ = *average* waiting time till the occurrence of the first point = μ^{-1}
- This is consistent with our assumption that the average time interval between points is $1/\mu$ seconds
- Let $P_k(t) = P\{\text{exactly k points in interval } (0,t]\}$
- $P_k(0) = P\{\text{k points in } (0,0]\} = P\{\text{k points in } \emptyset\} = 0$ for $k > 0$
- Set up and solve differential equation for $P_k(t)$
- Differential equation involves $P_{k-1}(t)$
- $P_k(t+\Delta t) = P\{\text{exactly k points have occurred upto time } t+\Delta t\}$



- $P_k(t+\Delta t) = P(AB)+P(CD)$ where $A = \{\text{k points upto time t}\}$; $B = \{\text{no points during } (t,t+\Delta t]\}$

  $C = \{\text{k–1 points upto time t}\}$; $D = \{\text{1 point during } (t,t+\Delta t]\}$
- $P_k(t+\Delta t) = P(AB)+P(CD) = P_k(t)(1-\mu\Delta t) + P_{k-1}(t)\mu\Delta t$

  $$\frac{P_k(t+\Delta t)-P_k(t)}{\Delta t} = -\mu P_k(t)+\mu P_{k-1}(t)$$
- $\frac{d}{dt}P_k(t) = -\mu P_k(t) + \mu P_{k-1}(t); \qquad P_k(0) = 0$
- Solve for $P_1(t)$, then for $P_2(t)$, and so on … till we find $P_k(t)$
- $\frac{d}{dt}P_1(t) = -\mu P_1(t) + \mu P_0(t) = -\mu P_1(t) + \mu\exp(-\mu t)$
- LaPlace transform solution
- $\frac{d}{dt}P_k(t) = -\mu P_k(t) + \mu P_{k-1}(t); \qquad P_k(0) = 0$

  $s\pounds\{P_k(t)\} = -\mu[\pounds\{P_k(t)\}-\pounds\{P_{k-1}(t)\}]$
- $\pounds\{P_k(t)\} = \frac{\mu}{s+\mu}\pounds\{P_{k-1}(t)\}$
- $\pounds\{P_k(t)\} = \frac{\mu}{s+\mu}\pounds\{P_{k-1}(t)\} = \left(\frac{\mu}{s+\mu}\right)^2\pounds\{P_{k-2}(t)\} = \left(\frac{\mu}{s+\mu}\right)^3\pounds\{P_{k-3}(t)\} = \ldots$

  $= \left(\frac{\mu}{s+\mu}\right)^k\pounds\{P_0(t)\}$

- $\pounds\{P_k(t)\} = \left(\frac{\mu}{s+\mu}\right)^k \pounds\{P_0(t)\} = \frac{\mu^k}{(s+\mu)^{k+1}}$ because $\pounds\{P_0(t)\} = \pounds\{\exp(-\mu t)\} = \frac{1}{s+\mu}$
- $P_k(t) = \exp(-\mu t)\frac{(\mu t)^k}{k!}$ for $t \geq 0$
- For k = 0, this gives $P_0(t)$ (if you remember that 0! = 1!)
- For k > 0, $P_k(0) = 0$; $P_k(t)$ is small for small t, increases to a maximum at $t = k/\mu$, and then decays away to 0
- k points can be expected to occur in roughly k/μ seconds,
- $P_k(t)$ is maximum at $t = k/\mu$
- For $t << k/\mu$, far fewer than k points <u>should</u> have occurred $P_k(t)$ is small for t small
- For $t >> k/\mu$, many more than k points <u>should</u> occur $P_k(t) \to 0$ as $t \to \infty$
- $\mathbf{X}_{k+1}$ = time of occurrence of the (k+1)th point
- Event $\{\mathbf{X}_{k+1} > u\}$ occurs iff no more than k points occur during the time interval (0,u], i.e. either 0 or 1 or 2 or … or k points occur during (0,u]
- For $u \geq 0$, $1 - F_{\mathbf{X}_{k+1}}(u) = P\{\mathbf{X}_{k+1} > u\} = P\{\leq k \text{ points}\} = P_0(u) + P_1(u) + \ldots P_k(u)$

  $= \exp(-\mu u)\left[1 + \frac{(\mu u)}{1!} + \ldots \frac{(\mu u)^k}{k!}\right]$
- $f_{\mathbf{X}_{k+1}}u) = -\frac{d}{du}\left[1 - F_{\mathbf{X}_{k+1}}(u)\right] = -\frac{d}{du}\exp(-\mu u)\left[1+\frac{(\mu u)}{1!}+\ldots+\frac{(\mu u)^k}{k!}\right]$

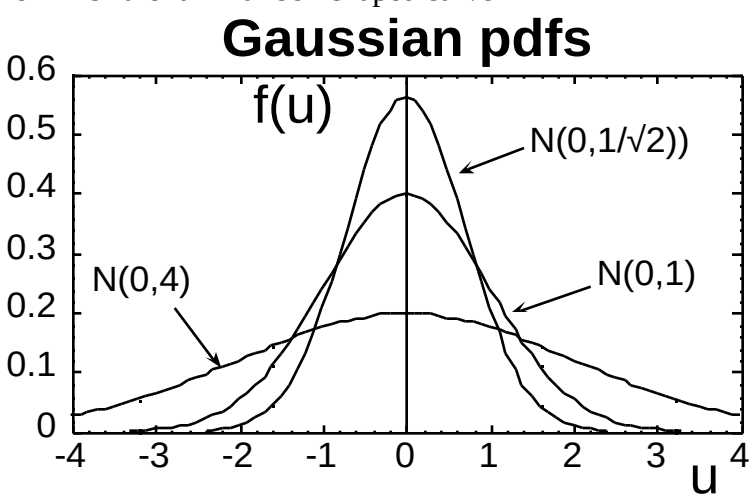
  To the tune of
  "This old man, he played one"
  Vee-dee-you (plus) you-dee-vee
  That's the "dee" of you-times-vee
  Remember, when the product rule you do,
  DON'T you say dee-vee dee-you!
- For $u \geq 0$, $f_{\mathbf{X}_{k+1}}(u) = -\frac{d}{du}\exp(-\mu u)\left[1+\frac{(\mu u)}{1!}+\ldots+\frac{(\mu u)^k}{k!}\right] = \mu\exp(-\mu u)\frac{(\mu u)^k}{k!}$

  !!!
- Verify that the terms cancel out to leave just the single one shown above
- For $u \geq 0$, $f_{\mathbf{X}_{k+1}}(u) = \mu\exp(-\mu u)\frac{(\mu u)^k}{k!}$
- $\mathbf{X}_{k+1}$ = time of occurrence of the (k+1)-th random point is a gamma random variable with parameters $(k+1,\mu)$
- $E[\mathbf{X}_k]$ = *average* waiting time till the occurrence of the k-th point $= \frac{k}{\mu}$
- This is consistent with our assumption that the average time interval between points is $1/\mu$
- Consider a fixed time interval (0,T]
- The number of random points occurring in this time interval is a <u>discrete</u> random variable **Y** taking on values 0, 1, 2, …
- What is $p_{\mathbf{Y}}(k)$, the pmf of **Y**?

- $p_{\mathbf{Y}}(k) = P\{\mathbf{Y} = k\} = P\{\text{exactly k points in } (0,T]\} = P_k(T) = \exp(-\mu T)\frac{(\mu T)^k}{k!}$
- For any fixed interval (0,T], $\mathbf{Y}$ is a Poisson random variable with parameter $\lambda = \mu T$
- $E[\mathbf{Y}] = \lambda = \mu T = \text{arrival rate} \times T$
- The number of points in $(t_1,t_2]$ is a Poisson random variable $\mathbf{Y}_1$ with parameter $\mu(t_2-t_1)$
- If $t_1 < t_2 \le t_3 < t_4$, and $\mathbf{Y}_1$ and $\mathbf{Y}_3$ are the numbers of points in $(t_1,t_2]$ and $(t_3,t_4]$, then these are *independent* Poisson RV
- For all m, n, $P\{\mathbf{Y}_1 = m, \mathbf{Y}_3 = n\} = P\{\mathbf{Y}_1 = m\}P\{\mathbf{Y}_3 = n\} = \exp(-\lambda_1)\frac{\lambda_1^m}{m!} \times \exp(-\lambda_3)\frac{\lambda_3^n}{n!}$

  Here, $\lambda_1 = \mu(t_2-t_1)$, $\lambda_3 = \mu(t_4-t_3)$
- Generalizes to multiple disjoint time intervals also
- Summary: Poisson process
- Points occur at random times
- μ points occur each second on *average*, and the *average interval* between successive points is $1/\mu$
- On average, μT points occur in any T second interval
- *Actual* number of points in T seconds is a Poisson *random* variable with parameter μT
- Numbers of points in *disjoint* time intervals of lengths T_1, T_2, … are independent Poisson random variables with parameters μT_1, μT_2, …
- $P\{\text{one point occurs in a } \textit{small} \text{ interval of length } \Delta t\} \approx \mu\Delta t$
- $P_k(T) = P\{\text{exactly k points in a T second interval}\} = P\{\mathbf{Y} = k\} = \exp(-\mu T)\frac{(\mu T)^k}{k!}$
- $\mathbf{X}_k$ = time of occurrence of the k-th point
- $\mathbf{X}_k$ is a gamma random variable with parameters (k,μ)
- $\mathbf{X}_1$ is an exponential random variable with parameter μ
- $\mathbf{X}_{k+i} - \mathbf{X}_i$ is a gamma random variable with parameters (k,μ)
- **Gaussian Random Variables**
- $\mathbf{X}$ is called a Gaussian (or normal) random variable with mean μ and variance σ^2 if its pdf is given by $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right]$ for all u, $-\infty < u < \infty$
- Notation: $\mathbf{X}$ is $N(\mu,\sigma^2)$
- The mean and variance of a Gaussian random variable specify its pdf completely
- If $\mathbf{X}$ is $N(3,25)$, its pdf is $f(u) = \frac{1}{5\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-3}{5}\right)^2\right]$ for all u, $-\infty < u < \infty$
- $\alpha\exp(-\beta u^2+\gamma u+\delta)$ is Gaussian

- $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right]$
- Read for yourself the proof that this is a valid pdf and that $E[\mathbf{X}] = \mu$, $\mathrm{var}(\mathbf{X}) = \sigma^2$
- pdf is symmetric about $u = \mu$, that is, $f(\mu+a) = f(\mu-a)$ median = μ
- $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right]$
- $(u-\mu)^2 \geq 0$ and hence exp[•] has maximum value 1 at $u = \mu$
- $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$ is the maximum value of the pdf mode = μ
- Mean = median = mode = μ
- The pdf of $\mathbf{X}$ is "the familiar bell-shaped curve"



- $\mathrm{var}(\mathbf{X}) = \sigma^2$ is a measure of the spread of the distribution
- The larger the variance, the more widely dispersed the mass
- Almost all of the probability mass lies inside the interval $(\mu-3\sigma,\ \mu+3\sigma)$
- ALWAYS write the Gaussian pdf as $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right]$
- NEVER, EVER write it as $f(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right]$
- Avoid creeping radicalism
- Gaussian random variables arise "normally" in studies of a very wide variety of natural phenomena
- The net effect produced by a very large number of small causes has a Gaussian distribution
- Central limit theorem
- **Examples:** Random motion of electrons in a conductor creates fluctuating noise voltages

- Small charge on electron small effect on the potential difference
- There are a large number of electrons
- Model of noise voltage as a Gaussian random variable allows for the possibility of arbitrarily large voltages (but these have very small probability)
- Kinetic theory of gases: the velocities of molecules are modeled as Gaussian random variables
- SAT and GRE scores are modeled as $N(500,100^2)$ Gaussian random variables (scores > 800 or < 200 are set to 800 and 200 respectively
- Measurement errors in experiments are modeled as Gaussian random variables: normal law of errors
- Henri Poincare´: Everybody believes in the normal law of errors: the physicists because they believe that it is a mathematical theorem, and the mathematicians because they believe that it is an experimentally observed fact
- An $N(0,1)$ random variable is called a *standard* or *unit* Gaussian random variable: its pdf is

  $\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right]$ Its CDF is denoted by $\Phi(u)$

- $\Phi(u)$ = area under pdf curve φ to the left of point u *cannot* be expressed in terms of simple functions
- Values of $\Phi(u)$ are tabulated
- Many scientific calculators have built-in routines to calculate $\Phi(u)$. Find out if yours can calculate $\Phi(u)$
- DO NOT attempt to integrate as follows:

  $\Phi(u) = \int_0^u \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right] du = \frac{1}{\sqrt{2\pi}u} \exp\left[-\frac{1}{2}u^2\right]$ This is sheer nonsense!

- The CDF of *any* $N(\mu,\sigma^2)$ random variable can be expressed in terms of $\Phi(u)$
- The *normalized distance* of any point t on the real line from the mean μ is defined as $(t-\mu)/\sigma$ = distance from mean in units of the standard deviation σ
- F(t) = area under pdf curve f(u) to the left of t = $\int_{-\infty}^{t} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right] du$

  Put $(u-\mu)/\sigma = v$, $du/\sigma = dv$ $u = -\infty$ $v = -\infty$ $u = t$ $v = (t-\mu)/\sigma$

- $F(t) = \int_{-\infty}^{(t-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}u^2\right] du$

- If $\mathbf{X}$ is $N(\mu,\sigma^2)$, then its CDF is given by $F(u) = \Phi\left(\frac{u-\mu}{\sigma}\right)$

- Note that the argument of Φ is just the normalized distance of u from μ

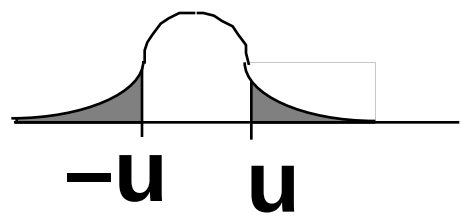
- The values of the CDF F(u) of a $N(\mu,\sigma^2)$ random variable can be found from tables of $\Phi(u)$ using $F(u) = \Phi((u-\mu)/\sigma)$
- Probability calculations for any $N(\mu,\sigma^2)$ random variable can be done using the tables for $\Phi(u)$
- **Example:** The $N(500,100^2)$ random variable **X** denotes the SAT score of a randomly chosen high-school student

  $P\{\mathbf{X} < 650\} = \Phi((650-500)/100) = \Phi(1.5) = 0.9332$

  $P\{\mathbf{X} > 700\} = 1 - \Phi(2) = 0.0228$

  $P\{\mathbf{X} < 400\} = \Phi(-1) = ??$

  $P\{\mathbf{X} < 400\} = \Phi(-1) = ??$
- Tables for $\Phi(\bullet)$ list values for positive arguments only



$\Phi(-u) = 1 - \Phi(u)$

$P\{\mathbf{X} < 400\} = \Phi(-1) = 1 - \Phi(1)$

- $P\{400 < \mathbf{X} < 600\} = \Phi(1) - \Phi(-1) = \Phi(1) - [1 - \Phi(1)] = 2\Phi(1) - 1$
- $\Phi(0) = ?$ $\Phi(0) =$
- Hence, $2\Phi(1) > 1$ and thus $2\Phi(1) - 1 > 0$
- Beware of miscalculations leading to $2\Phi(-1) - 1 < 0$
- $Q(u) = 1 - \Phi(u)$ = complementary CDF of $N(0,1)$ random variable

  $= \int_u^\infty \varphi(t)dt = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}t^2\right]dt$ decreases from 1 at $u = -\infty$ to 0 at $u = +\infty$
- For $u > 0$, $(u^{-1} - u^{-3})\varphi(u) < Q(u) < u^{-1}\varphi(u)$
  This bound is not useful for small values of u, but is very tight for large values of u
- For $u \geq 0$, $Q(u) \leq 0.5\exp(-u^2/2)$
  This bound is useful for small values of u but is looser than the one above for large u
- For *any* random variable **X** with mean μ and variance σ^2, let $\mathbf{Y} = (\mathbf{X}-\mu)/\sigma$
- $E[\mathbf{Y}] = E[(\mathbf{X}-\mu)/\sigma] = 0$
- $E[\mathbf{Y}^2] = E[((\mathbf{X}-\mu)/\sigma)^2]$ $\sigma^{-2}E[(\mathbf{X}-\mu)^2] = \sigma^{-2}\sigma^2 = 1$ so that $\mathrm{var}(\mathbf{Y}) = 1$
- If **X** is $N(\mu,\sigma^2)$, then $\mathbf{Y} = (\mathbf{X}-\mu)/\sigma$ is $N(0,1)$

- $P\{\mathbf{Y} < v\} = P\{(\mathbf{X}-\mu)/\sigma < v\} = P\{\mathbf{X} < \mu + v\sigma\} = F(\mu + v\sigma)$

  $= \Phi\left(\frac{\mu + v\sigma - \mu}{\sigma}\right) = \Phi(v)$ = CDF of a $N(0,1)$ random variable
- More generally, *any* linear transformation on a Gaussian random variable yields a Gaussian random variable
- This is true even when many Gaussian random variables are involved: linear transformations of Gaussian random variables yield Gaussian random variables
- Let $\mathbf{Y} = a\mathbf{X} + b$

  $P\{\mathbf{Y} < v\} = P\{a\mathbf{X} + b < v\} = P\{\mathbf{X} < (v-b)/a\} = \Phi\left(\frac{\frac{v-b}{a} - \mu}{\sigma}\right) = \Phi\left(\frac{v-a\mu-b}{a\sigma}\right)$
- If $\mathbf{X}$ is $N(\mu,\sigma^2)$, then $\mathbf{Y} = a\mathbf{X}+b$ is $N(a\mu+b,a^2\sigma^2)$
- For *any* random variable $\mathbf{X}$ with mean μ and variance σ^2, $\mathbf{Y} = a\mathbf{X}+b$ has mean

  $E[\mathbf{Y}] = E[a\mathbf{X}+b] = a\mu+b$ and $\mathrm{var}(\mathbf{Y}) = E[(\mathbf{Y}-a\mu-b)^2] = E[(a\mathbf{X}+b-a\mu-b)^2]$

  $= a^2E[(\mathbf{X}-\mu)^2] = a^2\sigma^2$
- For the *special* case when $\mathbf{X}$ is Gaussian, $\mathbf{Y} = a\mathbf{X}+b$ is Gaussian too
- To find the pdf of a Gaussian random variable, we need only find its mean and variance and then we can just *write* down the pdf
- $E[\mathbf{Y}] = E[a\mathbf{X}+b] = a\mu+b$ $\quad \mathrm{var}(\mathbf{Y}) = \mathrm{var}(a\mathbf{X}+b) = a^2\mathrm{var}(\mathbf{X})$ are general results that hold for *any* random variable
- For the case when $\mathbf{X}$ is Gaussian, so is $\mathbf{Y}$ and these calculations suffice to tell us the pdf of $\mathbf{Y}$ too
- $\mathbf{X}$ is $N(0,1)$ and $\mathbf{Y} = \mathbf{X}^2$
- $\mathbf{Y}$ takes on values ≥ 0 only
- For any $v \geq 0$, $F_{\mathbf{Y}}(v) = P\{\mathbf{Y} < v\} = P\{\mathbf{X}^2 < v\} = P\{-\sqrt{v} < \mathbf{X} < \sqrt{v}\} = \Phi(\sqrt{v}) - \Phi(-\sqrt{v})$
- $\frac{d}{du}\Phi(u) = \varphi(u)$
- $f_{\mathbf{Y}}(v) = \frac{1}{2\sqrt{v}}[\varphi(\sqrt{v})+\varphi(-\sqrt{v})] = \frac{1}{\sqrt{v}}\varphi(\sqrt{v}) = \frac{1}{\sqrt{2\pi v}}\exp(-v/2)$ = gamma pdf $t = 1/2$, $\lambda = 1/2$

  A.k.a. chi-squared (χ^2) pdf with one degree of freedom
- Gamma (n/2, 1/2) is a χ^2 pdf with n degrees of freedom
- **Hazard rates**
- System is put into operation at time $t = 0$ and fails at some later time $\mathbf{X}$
- The *time to failure* or the *lifetime* of the system is modeled as a continuous *random variable* $\mathbf{X} > 0$
- $\mathbf{X} > 0 \qquad F(u) = 0$ for $u \leq 0$
- $f(u) \to 0$ as $u \to \infty$
- Regardless of how the pdf f(u) behaves (increases, or decreases, or oscillates) for small values of u, it ultimately must decay away to 0
- For any u_0 such that $f(u_0) > 0$, we can find an $u_1 > u_0$ such that $f(u_0) > f(u)$ for all $u \geq u_1$

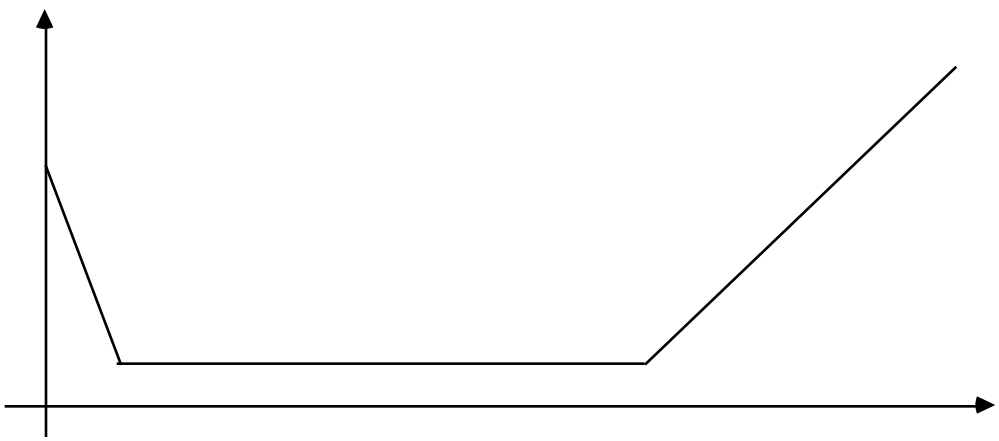
- P{system fails between time $t = u$ and $t = u+\delta$} = $P\{u < \mathbf{X} \le u+\delta\} \approx f(u)\times\delta \to 0$ as $u \to \infty$
- Suppose $u_0 = 21$, $u_1 = 80$
- $P\{21 < \mathbf{X} \le 21+\delta\}$ is larger than $P\{80 < \mathbf{X} \le 80+\delta\}$ and $P\{81 < \mathbf{X} \le 81+\delta\}$ and …?
- This result seems contrary to experience
- Experience indicates that older (more worn out) systems have a greater chance of failing than newer ones
- But the math says that your chances of dying in the next 24 hours are much greater today than when you will be 80 years old!
- What is the explanation? • "You should live so long"
- P{dying within next 24 hours} is of interest only to those alive right now
- $P\{u < \mathbf{X} \le u+\delta\}$ is less interesting than the conditional probability

$$P\{u < \mathbf{X} \le u+\delta \mid \mathbf{X} > u\} = \frac{P(\{u < \mathbf{X} \le u+\delta\} \cap \{\mathbf{X} > u\})}{P\{\mathbf{X} > u\}}$$
$$= \frac{P\{u < \mathbf{X} \le u+\delta\}}{P\{\mathbf{X} > u\}} \approx \frac{f(u)\times\delta}{1 - F(u)}$$

- The *hazard rate* function h(u), also called the *failure rate* function, of $\mathbf{X}$ is given by $h(u) = \dfrac{f(u)}{1 - F(u)}$
- P{system fails within the next δ seconds *given that it is working* at time $t = u$} $\approx h(u)\times\delta$
- $P\{u < \mathbf{X} \le u+\delta \mid \mathbf{X} > u\} \approx h(u)\times\delta$
- Given the CDF or the pdf of $\mathbf{X}$, it is easy to find the hazard rate function
- Find the *complementary* CDF from the pdf, or the pdf from the CDF, and set $h(u) = \dfrac{f(u)}{1 - F(u)}$
- **Example:** The pdf of $\mathbf{X}$ is $f(u) = \dfrac{u}{\sigma^2}\exp\left(-\dfrac{u^2}{2\sigma^2}\right)$ for $u > 0$
- For any *number* $t > 0$, $1 - F(t)$ = area under pdf to the right of $t = \int_t^\infty f(u)du$
- $1 - F(t) = \int_t^\infty \dfrac{u}{\sigma^2}\exp\left(-\dfrac{u^2}{2\sigma^2}\right)du = -\exp\left(-\dfrac{u^2}{2\sigma^2}\right)\Big|_t^\infty = \exp\left(-\dfrac{t^2}{2\sigma^2}\right)$ Hence, $h(u) = \dfrac{u}{\sigma^2}$ for $u > 0$
- $\mathbf{X}$ is called a Rayleigh random variable. Its hazard rate increases linearly with age
- $E[\mathbf{X}] = \sigma\sqrt{\pi/2}$ • $E[\mathbf{X}^2] = 2\sigma^2$
- The amplitude of the thermal noise at the output of a narrowband filter is usually modeled as a Rayleigh random variable
- **Example:** $\mathbf{X}$ has pdf $f(u) = \lambda\exp(-\lambda u)$ for $u > 0$
- $1 - F(t) = \int_t^\infty \lambda\exp(-\lambda u)du = \exp(-\lambda t)$ • $h(u) = \lambda$ for all $u > 0$

- Constant hazard rate means the system does not wear out
- Semiconductor devices seem to have constant hazard rates
- Given the pdf or CDF, it is easy to find the hazard rate
- The hazard rate also uniquely specifies the pdf and CDF
- $h(u) = \dfrac{f(u)}{1 - F(u)}$ for all $u > 0$
- Integrals should be equal too!
- For any real number t, $\int_0^t h(u)du = \int_0^t \frac{f(u)}{1 - F(u)}du = -\ln(1-F(u))\Big|_0^t = -\ln(1 - F(t))$

  since $F(0) = 0$
- $F(t) = 1-\exp\left(-\int_0^t h(u)du\right)$ for $t > 0$
- A typical hazard rate function

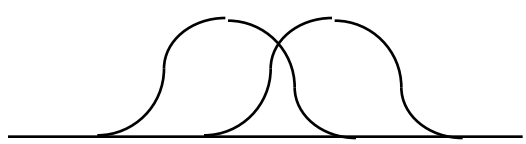


- **Semiconductor devices are still in the flat part of the curve**
- **High-reliability applications require a burn-in period to weed out the lemons**
- Contrary to popular belief, hazard rate functions are NOT pdfs and they are NOT the probability of anything
- P{system fails within the next δ seconds *given that it is working* at time $t = u$} $\approx h(u)\times\delta$
- Must multiply by δ to get a probability
- Area under pdf curve = 1
- Area under hazard rate curve is infinite
- $F(t) = 1-\exp\left(-\int_0^t h(u)du\right)$ for $t > 0$
- $F(\infty) = 1$ and thus $\int_0^\infty h(u)du = \infty$
- If $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively have hazard rates $h_1(u)$ and $h_2(u)$ where $h_1(u) < h_2(u)$ for all u, then $E[\mathbf{X}_1] > E[\mathbf{X}_2]$
- For any t, there is more area from 0 upto t under h_2 than under h_1. Why? $h_1(u) < h_2(u)$

- $1-F_i(t) = \exp\left(-\int_0^t h_i(u)du\right),\ i = 1,2$
- $h_1 < h_2 \qquad 1 - F_1(t) > 1 - F_2(t)$ for all $t > 0$
- But, $E[\mathbf{X}_i] = \int_0^\infty 1 - F_i(t)\,dt.$ Hence, $E[\mathbf{X}_1] > E[\mathbf{X}_2]$

CONDITIONAL DISTRIBUTIONS

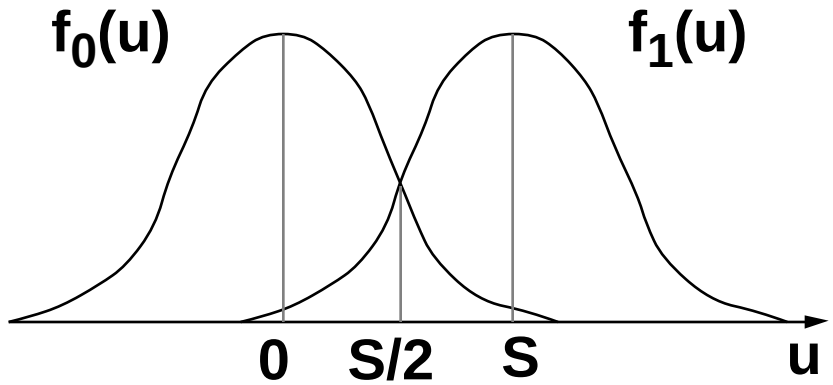
- Let A be an event of nonzero probability
- The conditional CDF of $\mathbf{X}$ given A is $F_{\mathbf{X}|A}(u|A) = P(\{\mathbf{X} \le u\}|A) = \dfrac{P(\{\mathbf{X} \le u\} \cap A)}{P(A)}$
- A conditional CDF has all the usual properties of CDFs
- The unconditional CDF can be found by combining the conditional CDFs using the theorem of total probability
- $P\{\mathbf{X} \le u\} = P\{\mathbf{X} \le u|A)P(A) + P\{\mathbf{X} \le u|A^c)P(A^c)$
- $F_{\mathbf{X}}(u) = F_{\mathbf{X}|A}(u|A)P(A) + F_{\mathbf{X}|A^c}(u|A^c)P(A^c)$
- The conditional pdf of $\mathbf{X}$ given A is the derivative of the conditional CDF
- $f_{\mathbf{X}|A}(u|A) = \dfrac{d}{du}F_{\mathbf{X}|A}(u|A)$
- $f_{\mathbf{X}}(u) = f_{\mathbf{X}|A}(u|A)P(A) + f_{\mathbf{X}|A^c}(u|A^c)P(A^c)$
- If $\mathbf{X}$ is a discrete random variable, then the above applies to conditional pmfs
- $p_{\mathbf{X}}(u) = p_{\mathbf{X}|A}(u|A)P(A) + p_{\mathbf{X}|A^c}(u|A^c)P(A^c)$
- **Example:** $P(A) = P(A^c) = 1/2$. The *conditional* pdf of $\mathbf{X}$ given A is $N(+1, 1)$ while the *conditional* pdf of $\mathbf{X}$ given A^c is $N(0, 1)$



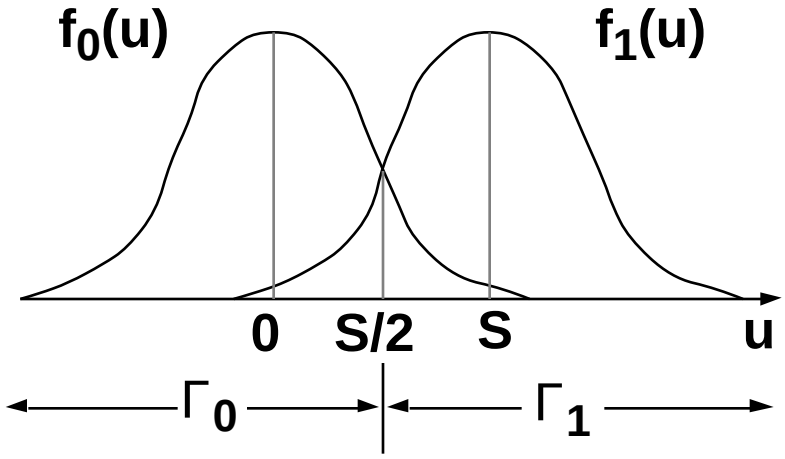
- $f_{\mathbf{X}|A}(u|A) = \varphi(u-1) \qquad f_{\mathbf{X}|A^c}(u|A^c) = \varphi(u)$
- The *unconditional* pdf is $f_{\mathbf{X}}(u) = \dfrac{\varphi(u-1) + \varphi(u)}{2}$
- $\mathbf{X}$ is *not* a Gaussian random variable
- Its unconditional pdf has two bumps at 0 and 1 respectively
- However, the *conditional* pdf of $\mathbf{X}$ is Gaussian when A occurs as well as when A^c occurs
- $\mathbf{X}$ is a *conditionally* Gaussian random variable
- **Example:** $\mathbf{X}$ is the receiver output in a radar system. If the target is absent, $\mathbf{X}$ is an $N(0, 1)$ random variable. If the target is present, an echo of amplitude S is also present and $\mathbf{X}$ is an $N(S, 1)$ random variable
- We have two *hypotheses*

- H_0 = target absent
- H_1 = target present
- *Conditioned* on H_0, $\mathbf{X}$ is an $N(0, 1)$ random variable
  Conditioned on H_1, $\mathbf{X}$ is an $N(S, 1)$ random variable
- **Notation:**
  $f_{\mathbf{X}|H_0}(u|H_0)$ is denoted by $f_0(u)$: $f_0(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2) = \varphi(u)$
  $f_{\mathbf{X}|H_1}(u|H_1)$ is denoted by $f_1(u)$: $f_1(u) = \frac{1}{\sqrt{2\pi}}\exp(-(u-S)^2/2) = \varphi(u-S)$
- H_0: $\mathbf{X} = \mathbf{N}$ $\mathbf{X}$ is $N(0, 1)$
  H_1: $\mathbf{X} = S + \mathbf{N}$ $\mathbf{X}$ is $N(S, 1)$ where $\mathbf{N}$ is $N(0, 1)$
  Reality check: Linear function of Gaussian is Gaussian with $E[S + \mathbf{N}] = S + E[\mathbf{N}] = S$ and $\mathrm{var}(S + \mathbf{N}) = \mathrm{var}(\mathbf{N}) = 1$
- The radar system has no idea which of the two hypotheses is true. If the receiver output $\mathbf{X}$ has taken on value 2.3678, is a target present or not?
- This is a statistical decision problem or a hypothesis testing problem
- Maximum-likelihood decision rule
- Choose the hypothesis that maximizes the probability of the observation
- But, since $\mathbf{X}$ is continuous, $P(\{\mathbf{X} = 2.3678\}|H_i) = 0$ both for i = 0 and for i = 1, so the rule is of no help!
- $\mathbf{X} = 2.3678$ often means that $\mathbf{X}$ has value between 2.36775 and 2.36785
- More generally, if we observe $\mathbf{X} = u$, we can compare
  $P(\{u-\delta < \mathbf{X} < u+\delta\}|H_0)$ and $P(\{u-\delta < \mathbf{X} < u+\delta\}|H_1)$ in the limit as $\delta \to 0$
- If H_0 is the true hypothesis, i.e. no target is present, $P(\{2.36775 < \mathbf{X} < 2.36785\}|H_0)$
  $\approx f_0(2.3678)\times 0.0001$
- If H_1 is the true hypothesis, i.e. a target is present, $P(\{2.36775 < \mathbf{X} < 2.36785\}|H_1)$
  $\approx f_1(2.3678)\times 0.0001$
- To determine which probability is larger, we can just compare the values of $f_0(2.3678)$ and $f_1(2.3678)$
- **Maximum-likelihood decision rule:** If $\mathbf{X}$ has taken on value u, then choose H_0 or H_1 according as $f_0(u)$ or $f_1(u)$ is larger
- The likelihood ratio is defined to be $\Lambda = \frac{f_1(u)}{f_0(u)}$
- $\Lambda = \frac{f_1(u)}{f_0(u)} = \frac{\exp(-(u-S)^2/2)}{\exp(-u^2/2)} = \exp(-S^2/2+uS) \underset{H_0}{\overset{H_1}{\gtrless}} 1?$
  $\Lambda > 1$ if $-S^2/2+uS > 0$ $u > S/2$
  $\Lambda < 1$ if $-S^2/2+uS < 0$ $u < S/2$
- We thus have a *threshold* test

- It is not necessary to actually compute Λ; we can simply compare the observation to a threshold of S/2 and decide
- Compare the receiver output $\mathbf{X}$ to S/2. If output is smaller than S/2, declare that there is no target present (i.e. choose hypothesis H_0) while if the output is larger than S/2, declare that there is a target present, (i.e. choose hypothesis H_1)
- Sketch $f_1(u)$ and $f_0(u)$ on same axes

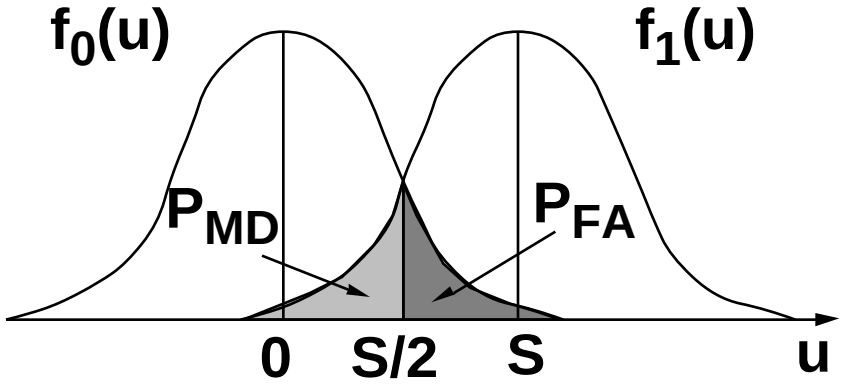


- One pdf cannot dominate another
- $\Gamma_0 = \{u : f_0(u) > f_1(u)\}$
- $\Gamma_1 = \{u : f_1(u) > f_0(u)\}$
- If $\mathbf{X}$ has taken on value α, decide in favor of H_0 if $\alpha \quad \Gamma_0$ and in favor of H_1 if $\alpha \quad \Gamma_1$
- In our radar example, $\Gamma_0 = (-\infty, S/2)$; $\Gamma_1 = (S/2, \infty)$

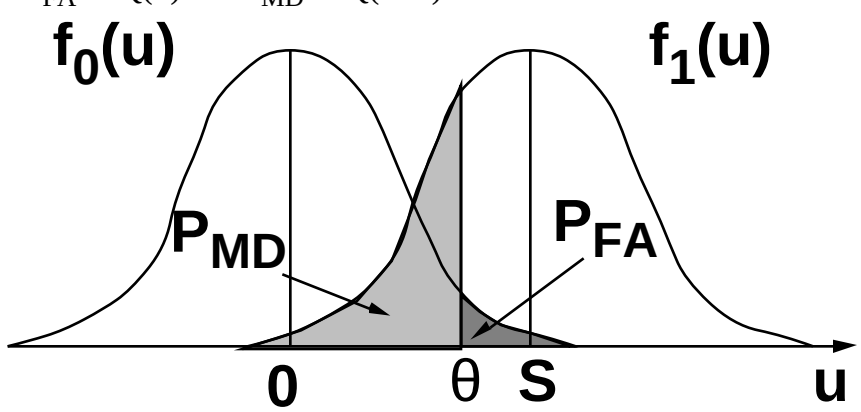


- Error probabilities: P_{FA} = P(false alarm), P(missed detection) = P_{MD}
- A false alarm occurs when H_0 is true but $\mathbf{X} \quad \Gamma_1$
- When H_0 is true, the conditional pdf of $\mathbf{X}$ is $f_0(u)$: $P_{FA} = P\{\mathbf{X} \quad \Gamma_1 | H_0\} = \int_{u \quad \Gamma_1} f_0(u)du$
- Detection is missed when H_1 is true but $\mathbf{X} \quad \Gamma_0$

- When H_1 is true, the conditional pdf of $\mathbf{X}$ is $f_1(u)$: $P_{MD} = P\{\mathbf{X} \quad \Gamma_0 | H_1\} = \int_{\Gamma_0} f_1(u)du$ u
- For our radar example,



- $P_{FA} = P_{MD}$ (by symmetry!) = $\int_{S/2}^{\infty} \varphi(u)du = Q(S/2)$
- If the observed value is compared to an arbitrarily chosen threshold θ (instead of S/2 as above), then $P_{FA} = Q(\theta)$ and $P_{MD} = Q(S-\theta)$



- $P_{FA} = Q(\theta)$ and $P_{MD} = Q(S-\theta)$ where Q(x) is the complementary CDF
- $P_{FA} = Q(\theta)$ decreases from 1 to 0 as θ increases from $-\infty$ to $+\infty$
- $P_{MD} = Q(S-\theta)$ increases from 0 to 1 as θ increases from $-\infty$ to $+\infty$
- Choosing $\theta = S/2$ equalizes these probabilities
- **Bayesian decision rules**
- With an arbitrary threshold θ, $P_{FA} = Q(\theta)$ and $P_{MD} = Q(S-\theta)$
- When H_0 is true, we decide incorrectly that H_1 is true with probability P_{FA}
- Let C_0 denote the penalty for making this wrong decision
- When H_1 is true, we decide incorrectly that H_0 is true with probability P_{MD}
- Let C_1 denote the penalty for making this wrong decision

- Let $\pi_0 = P(H_0 \text{ is true})$
  $\pi_1 = 1 - \pi_0 = P(H_1 \text{ is true})$
- The *average* penalty (also called the *average risk*) is $\pi_0 C_0 P_{FA} + \pi_1 C_1 P_{MD}$
  $= \pi_0 C_0 Q(\theta) + \pi_1 C_1 Q(S-\theta)$
- For the special case $C_0 = C_1 = 1$, the average risk is simply the*average error probability*
- We can minimize the average risk by clever choice of θ. This is the Bayes threshold
- Standard calculus technique:
  Differentiate the average risk $\pi_0 C_0 Q(\theta) + \pi_1 C_1 Q(S-\theta)$ with respect to θ and equate to 0
- $Q(u) = 1 - \Phi(u)$; $\frac{d}{du}\Phi(u) = \varphi(u)$
- $\frac{d}{du}\pi_0 C_0 Q(\theta) + \pi_1 C_1 Q(S-\theta) = -\pi_0 C_0 \varphi(\theta) + \pi_1 C_1 \varphi(S-\theta) = 0$
- $-\pi_0 C_0 + \pi_1 C_1 \exp(-\frac{S^2}{2} + S\theta) = 0$
- $\theta_{Bayes} = \frac{S}{2} + \frac{1}{S}\ln\left(\frac{\pi_0 C_0}{\pi_1 C_1}\right) = \theta_{ML} + \gamma$ $\qquad \gamma$ = bias term
- Bayes risk = average risk when the threshold is set to θ_{Bayes} is given by
  $\pi_0 C_0 Q(S/2+\gamma) + \pi_1 C_1 Q(S/2-\gamma)$
- With any other threshold setting, you will pay a larger penalty (on average)
- If $\pi_0 C_0 > \pi_1 C_1$, $\gamma > 0$ i.e., we increase the threshold to something larger than S/2 in an effort to reduce P_{FA}
- If S is large, costs have only a small effect on the threshold
- **Neyman-Pearson decision rule**

  We want P_{FA} to be no larger than α, and to have $1 - P_{MD}$ = power of the test as large as possible
- Of the myriad decision rules that satisfy $P_{FA} \le \alpha$, the rule that also maximizes $1 - P_{MD}$ is the Neyman-Pearson decision rule
- Let θ_{NP} be the solution to $Q(\theta_{NP}) = \alpha$
- For $0 < \alpha \le 0.5$, $\theta_{NP} \approx \sqrt{-2 \cdot \ln \alpha}$ For a more exact solution, consult Abramowitz and Stegun,"Handbook of Mathematical Functions" Chapter 26
- Then, $P_{FA} = \alpha$; $P_{MD} = Q(S-\theta_{NP})$
- If θ is chosen to be any larger than θ_{NP}, then $P_{FA} < \alpha$ but P_{MD} is larger than $Q(S-\theta_{NP})$
- If θ is smaller, $P_{FA} > \alpha$