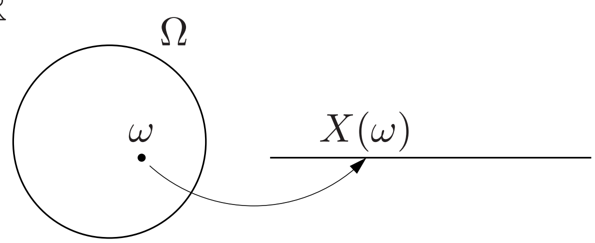
Lecture Notes 2
Random Variables

- Definition
- Discrete Random Variables: Probability mass function (pmf)
- Continuous Random Variables: Probability density function (pdf)
- Mean and Variance
- Cumulative Distribution Function (cdf)
- Functions of Random Variables

Corresponding pages from B&T textbook: 72–83, 86, 88, 90, 140–144, 146–150, 152–157, 179–186.

Random Variable

- A random variable is a real-valued variable that takes on values randomly
  Sounds nice, but not terribly precise or useful
- Mathematically, a *random variable* (r.v.) X is a real-valued function $X(\omega)$ over the sample space Ω of a random experiment, i.e., $X : \Omega \to \mathbb{R}$



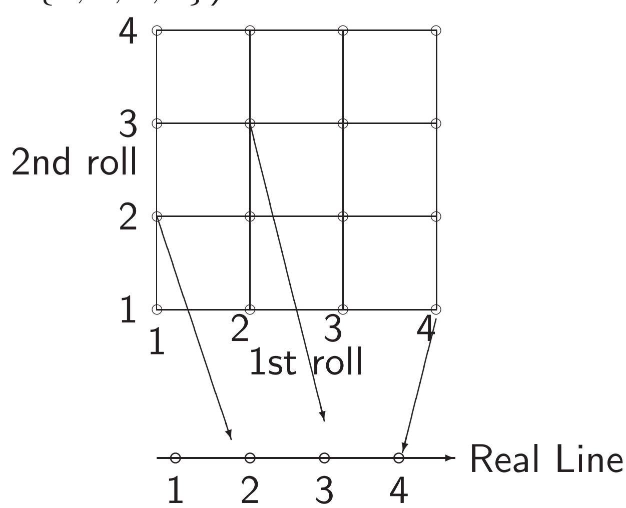
- Randomness comes from the fact that outcomes are random ($X(\omega)$ is a deterministic function)
- Notation:

- ◦ Always use upper case letters for random variables (X, Y, ...)
- ◦ Always use lower case letters for values of random variables: $X = x$ means that the random variable X takes on the value x

- Examples:

1. Flip a coin n times. Here $\Omega = \{H, T\}^n$. Define the random variable $X \in \{0, 1, 2, \ldots, n\}$ to be the number of heads
2. Roll a 4-sided die twice.

(a) Define the random variable X as the maximum of the two rolls ($X \in \{1, 2, 3, 4\}$)



(b) Define the random variable Y to be the sum of the outcomes of the two rolls ($Y \in \{2, 3, \cdots, 8\}$)

(c) Define the random variable Z to be 0 if the sum of the two rolls is odd and 1 if it is even

3. Flip coin until first heads shows up. Define the random variable $X \in \{1, 2. \ldots\}$ to be the number of flips until the first heads

4. Let $\Omega = \mathbb{R}$. Define the two random variables

(a) $X = \omega$

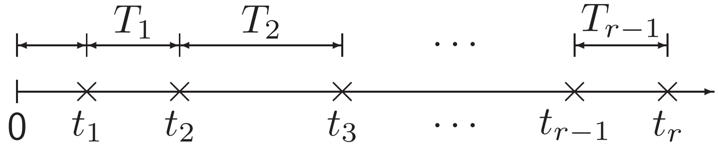
(b) $Y = \begin{cases} +1 & \text{for } \omega \geq 0 \\ -1 & \text{otherwise} \end{cases}$

5. n packets arrive at a node in a communication network. Here Ω is the set of arrival time sequences $(t_1, t_2, \ldots, t_n) \in (0, \infty)^n$

(a) Define the random variable N to be the number of packets arriving in the interval $(0, 1]$

(b) Define the random variable T to be the first interarrival time $(t_2 - t_1)$.
More generally, $T_N = N$th interarrival time $t_{N+1} - t_N$.
Note:

$$t_N = t_1 + \sum_{k=1}^{N-1} T_k$$

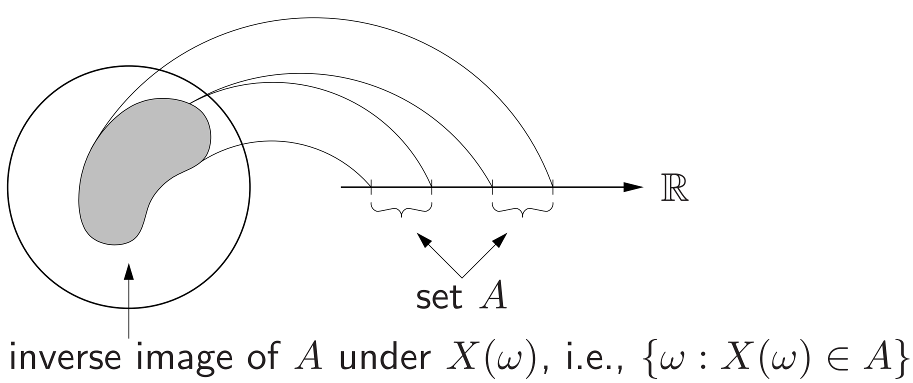


Arrivals and interarrivals important in many situations: Phone calls, ships arriving, orders being placed, patients signing in, queuing theory, . . .

- Why do we need random variables?
  - Random variable can represent the gain or loss in a random experiment, e.g., stock market
  - Random variable can represent a measurement over a random experiment, e.g., noise voltage on a resistor
- In most applications we care more about these costs/measurements than the underlying probability space
- Very often we work directly with random variables without knowing (or caring to know) the underlying probability space

Specifying a Random Variable

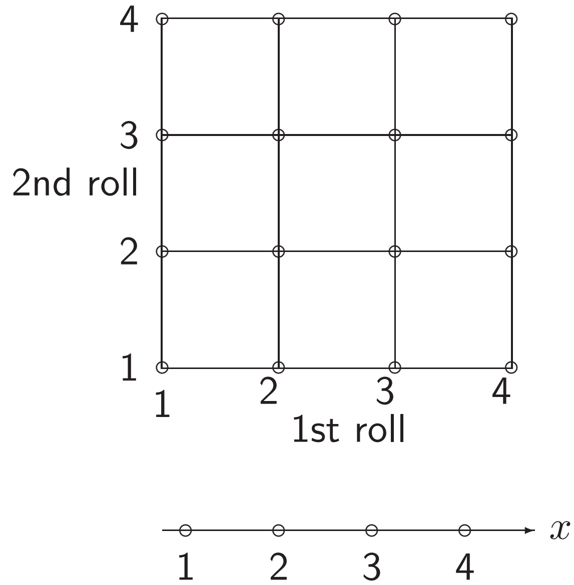
- *Specifying* a random variable means being able to determine the probability that $X \in A$ for any event $A \subset \mathbb{R}$, e.g., any interval
- To do so, we consider the *inverse image* of the set A under $X(\omega)$, $X^{-1}(A) = \{\omega : X(\omega) \in A\}$



- So, $X \in A$ iff $\omega \in \{\omega : X(\omega) \in A\}$, thus $\mathrm{P}(\{X \in A\}) = \mathrm{P}(\{w : X(\omega) \in A\})$, or in short

$$\boxed{\mathrm{P}\{X \in A\} = \mathrm{P}\{\omega : X(\omega) \in A\}}$$

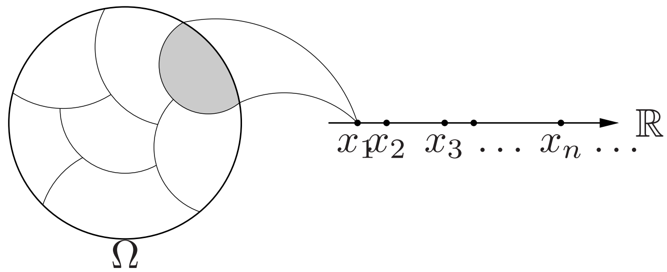
- Example: Roll fair 4-sided die twice independently: Define the r.v. X to be the maximum of the two rolls. What is the $\mathrm{P}\{0.5 < X \le 2\}$?



- We classify r.v.s as:
  - Discrete: X can assume only one of a countable number of values. Such r.v. can be specified by a *probability mass function* (pmf). Examples 1, 2, 3, 4(b), and 5(a) are of discrete r.v.s
  - Continuous: X can assume one of a continuum of values and the probability of each value is 0. Such r.v. can be specified by a *probability density function* (pdf). Examples 4(a) and 5(b) are of continuous r.v.s
  - Mixed: X is neither discrete nor continuous. Such r.v. (as well as discrete and continuous r.v.s) can be specified by a *cumulative distribution function* (cdf)

Discrete Random Variables

- A random variable is said to be *discrete* if for some countable set $\mathcal{X} \subset \mathbb{R}$, i.e., $\mathcal{X} = \{x_1, x_2, \ldots\}$, $\mathrm{P}\{X \in \mathcal{X}\} = 1$
- Examples 1, 2, 3, 4(b), and 5(a) are discrete random variables
- Here $X(\omega)$ partitions Ω into the sets $\{\omega : X(\omega) = x_i\}$ for $i = 1, 2, \ldots$. Therefore, to specify X, it suffices to know $\mathrm{P}\{X = x_i\}$ for all i

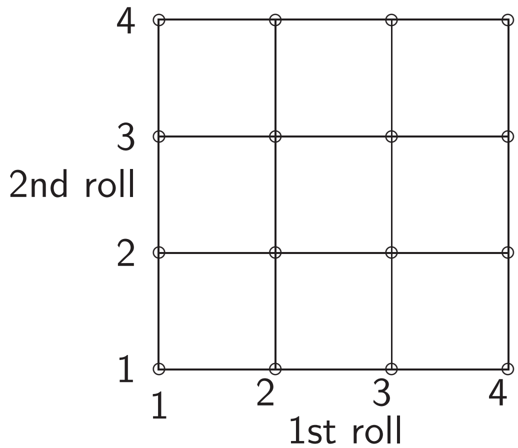


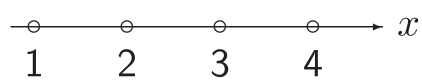
- A discrete random variable is thus completely specified by its *probability mass function* (pmf)

$$p_X(x) = \mathrm{P}\{X = x\} \text{ for all } x \in \mathcal{X}$$

- Clearly $p_X(x) \geq 0$ and $\sum_{x \in \mathcal{X}} p_X(x) = 1$

- Example: Roll a fair 4-sided die twice independently: Define the r.v. X to be the maximum of the two rolls





$p_X(x)$:

- Note that $p_X(x)$ can be viewed as a probability measure over a discrete sample space (even though the original sample space may be continuous as in examples 4(b) and 5(a))
- The probability of any event $A \subset \mathbb{R}$ is given by

$$\mathrm{P}\{X \in A\} = \sum_{x \in A \cap \mathcal{X}} p_X(x)$$

For the previous example $\mathrm{P}\{1 < X \leq 2.5 \text{ or } X \geq 3.5\} =$

- Notation: We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to mean that the discrete random variable X has pmf $p_X(x)$ or $p(x)$

Famous Probability Mass Functions

- *Uniform*: $X \sim \mathrm{Unif}(N)$ for N a positive integer has pmf

$$p_X(x) = \frac{1}{N}, x = 0, 1, 2, \ldots, N-1$$

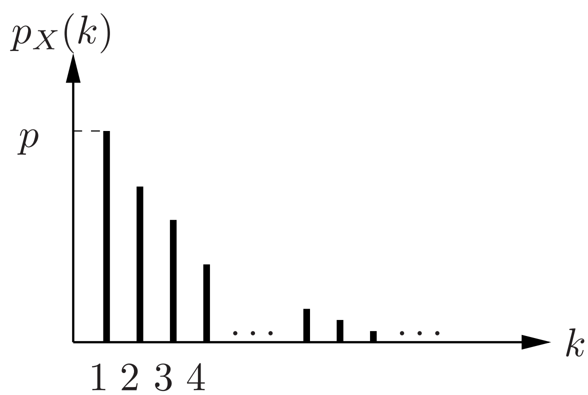
(Or any finite set with N elements. Fair coin $N = 1$, fair die $N = 6$, roulette wheel $N = 37$ (38 in the USA))

- *Bernoulli*: $X \sim \mathrm{Bern}(p)$ for $0 \le p \le 1$ has pmf

$$p_X(1) = p, \text{ and } p_X(0) = 1 - p$$

- *Geometric*: $X \sim \mathrm{Geom}(p)$ for $0 \le p \le 1$ has pmf

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



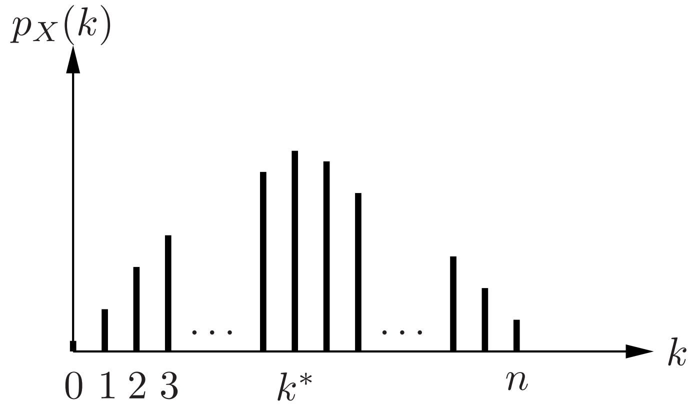
This r.v. represents, for example, the number of coin flips until the first heads shows up (assuming independent coin flips)

			k
H	$\longrightarrow$	1	p
TH	$\longrightarrow$	2	$(1-p)p$
TTH	$\longrightarrow$	3	$(1-p)^2 p$
TTTH	$\longrightarrow$	4	$(1-p)^3 p$
$\vdots$			

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

- *Binomial*: $X \sim \mathrm{B}(n,p)$ for integer $n > 0$ and $0 \le p \le 1$ has pmf

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



The maximum of $p_X(k)$ is attained at

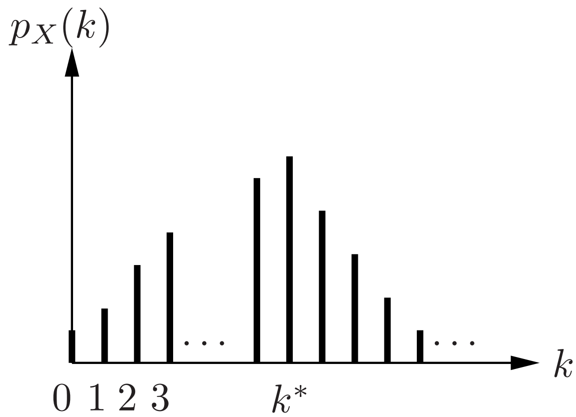
$$k^* = \begin{cases} (n+1)p,\ (n+1)p-1, & \text{if } (n+1)p \text{ is an integer} \\ \underbrace{[(n+1)p]}_{\text{integer part}}, & \text{otherwise} \end{cases}$$

The binomial r.v. represents, for example, the number of heads in n independent coin flips (see page 72 of Lecture Notes 1)

Flip coin with bias p until first head: (H=1, T=0)

- *Poisson*: $X \sim \mathrm{Poisson}(\lambda)$ for $\lambda > 0$ (called the *rate*) has pmf

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for } k = 0, 1, 2, \ldots$$



The maximum of $p_X(k)$ attained at

$$k^* = \begin{cases} \lambda,\ \lambda - 1, & \text{if } \lambda \text{ is an integer} \\ [\lambda], & \text{otherwise} \end{cases}$$

The Poisson r.v. often represents the number of random events, e.g., arrivals of packets, photons, customers, etc.in some time interval, e.g., $[0, 1)$

Common model (with origins in physics): Fix a time $t > 0$, then

$$P\{k \text{ arrivals in } [0, t)\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Note:

- If t very small, say Δt then $P\{1 \text{ arrival in } [0, \Delta t]\} \approx \lambda \Delta t$, the probability of more than one arrival in $[0, \Delta t)$ is negligible (higher order of t), and $P\{0 \text{ arrivals in } [0, \Delta t]\} \approx 1 - \lambda \Delta t$. Reasonable, e.g., for random phone calls or photons.
- Combined with an assumption of independence in nonoverlapping time intervals, this leads to powerful and accurate models of many random phenomena.

- Historically: Invented by Poisson as an approximation to the Binomial for large n with $p \propto 1/n$ based on an important property of e

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

To be precise: Poisson is the limit of Binomial when $p \propto \frac{1}{n}$, as $n \to \infty$

To show this let $X_n \sim \mathrm{B}(n, \frac{\lambda}{n})$ for $\lambda > 0$. For any fixed nonnegative integer k,

$$\begin{aligned}
p_{X_n}(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{(n-k)} \\
&= \frac{n(n-1)\ldots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(\frac{n-\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\ldots(n-k+1)}{(n-\lambda)^k}\frac{\lambda^k}{k!}\left(\frac{n-\lambda}{n}\right)^{n} \\
&= \frac{n(n-1)\ldots(n-k+1)}{(n-\lambda)(n-\lambda)\ldots(n-\lambda)}\frac{\lambda^k}{k!}\underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{\to e^{-\lambda}} \\
&\to \frac{\lambda^k}{k!}e^{-\lambda} \text{ as } n \to \infty
\end{aligned}$$

Continuous Random Variables

- Suppose a r.v. X can take on a continuum of values each with probability 0

  Examples:

  - Pick a number between 0 and 1
  - Measure the voltage across a heated resistor
  - Measure the phase of a random sinusoid . . .

- How do we describe probabilities of interesting events?
- Idea: For discrete r.v., we sum a pmf over points in a set to find its probability. For continuous r.v., integrate a probability *density* over a set to find its probability — analogous to mass density in physics (integrate mass density to get the mass)

Probability Density Function

- A continuous r.v. X can be specified by a *probability density function* $f_X(x)$ (pdf) such that, for any event A,

$$\mathrm{P}\{X \in A\} = \int_A f_X(x)\,dx$$

  For example, for $A = (a, b]$, the probability can be computed as

$$\mathrm{P}\{X \in (a, b]\} = \int_a^b f_X(x)\,dx$$

- Properties of $f_X(x)$:
  1. $f_X(x) \geq 0$
  2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

- Important note: $f_X(x)$ should not be interpreted as the probability that $X = x$, in fact it is *not* a probability measure, e.g., it can be > 1
- Can relate $f_X(x)$ to a probability using mean value theorem for integrals: Fix x and some $\Delta x > 0$. Then provided f_X is continuous over $(x, x + \Delta x]$,

$$\begin{aligned} \mathrm{P}\{X \in (x, x+\Delta x]\} &= \int_x^{x+\Delta x} f_X(\alpha)\, d\alpha \\ &= f_X(c)\, \Delta x \text{ for some } x \le c \le x + \Delta x \end{aligned}$$

Now, if Δx is sufficently small, then

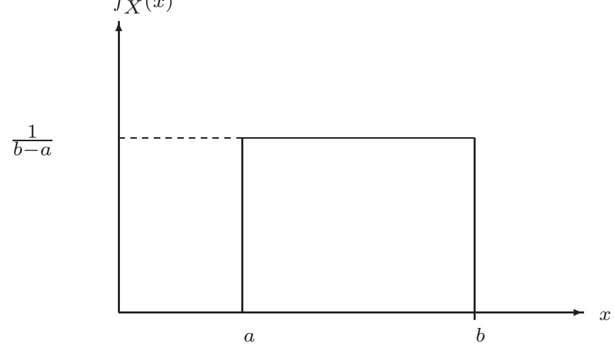
$$\mathrm{P}\{X \in (x, x+\Delta x]\} \approx f_X(x)\Delta x$$

- Notation: $X \sim f_X(x)$ means that X has pdf $f_X(x)$

Famous Probability Density Functions

- *Uniform*: $X \sim \mathrm{U}[a, b]$ for $b > a$ has the pdf

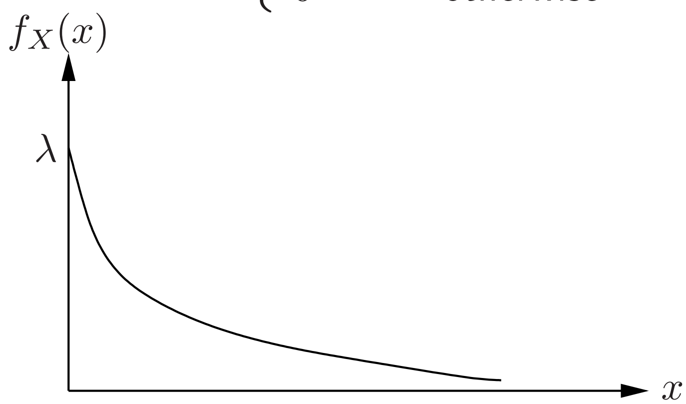
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



Uniform r.v. is commonly used to model quantization noise and finite precision computation error (roundoff error)

- *Exponential*: $X \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$ has the pdf

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



Exponential r.v. is commonly used to model *interarrival time* in a queue, i.e., time between two consecutive packet or customer arrivals, service time in a queue, and lifetime of a particle, time between busses, etc.

Example: Let $X \sim \mathrm{Exp}(0.1)$ be the customer service time at a bank (in minutes). The person ahead of you has been served for 5 minutes. What is the probability that you will wait another 5 minutes or more before getting served?

We want to find $\mathrm{P}\{X > 10 \,|\, X > 5\}$

Solution: By definition of conditional probability

$$\begin{aligned} \mathrm{P}\{X > 10 \,|\, X > 5\} &= \frac{\mathrm{P}\{X > 10,\ X > 5\}}{\mathrm{P}\{X > 5\}} \\ &= \frac{\mathrm{P}\{X > 10\}}{\mathrm{P}\{X > 5\}} \\ &= \frac{\int_{10}^{\infty} 0.1 e^{-0.1x}\, dx}{\int_{5}^{\infty} 0.1 e^{-0.1x}\, dx} \\ &= \frac{e^{-1}}{e^{-0.5}} = e^{-0.5} = \mathrm{P}\{X > 5\} \Rightarrow \end{aligned}$$

the conditional probability of waiting more than 5 additional minutes given that you have already waited more than 5 minutes is the same as the *un*conditional probability of waiting more than 5 minutes!

This is because the exponential r.v. is *memoryless*, which in general means that for any $0 \le x_1 < x$,

$$\mathrm{P}\{X > x \,|\, X > x_1\} = \mathrm{P}\{X > x - x_1\}$$

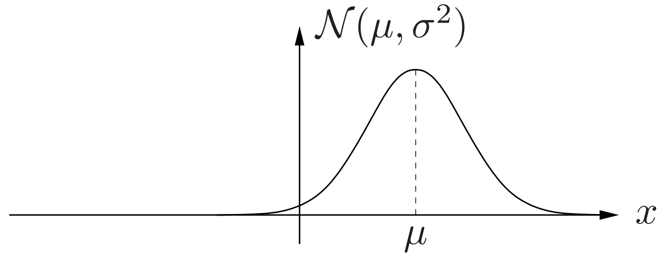
To show this, consider

$$\begin{aligned}
\mathrm{P}\{X > x \,|\, X > x_1\} &= \frac{\mathrm{P}\{X > x,\ X > x_1\}}{\mathrm{P}\{X > x_1\}} \\
&= \frac{\mathrm{P}\{X > x\}}{\mathrm{P}\{X > x_1\}} \text{ assuming } x > x_1 \\
&= \frac{\int_x^\infty \lambda e^{-\lambda x}\, dx}{\int_{x_1}^\infty \lambda e^{-\lambda x}\, dx} \\
&= \frac{e^{-\lambda x}}{e^{-\lambda x_1}} = e^{-\lambda(x - x_1)} \\
&= \mathrm{P}\{X > x - x_1\}
\end{aligned}$$

- *Gaussian (normal)*: $X \sim \mathcal{N}(\mu, \sigma^2)$ has pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for } -\infty < x < \infty,$$

where μ is the *mean* and σ^2 is the *variance*



Gaussian r.v.s are frequently encountered in nature, e.g., thermal and shot noise in electronic devices are Gaussian, and very frequently used in modelling various social, biological, and other phenomena

Mean and Variance

- A discrete (continuous) r.v. is *completely* specified by its pmf (pdf)
- It is often desirable to *summarize* the r.v. or *predict* its outcome in terms of one or a few numbers. What do we *expect* the value of the r.v. to be? What range of values around the mean do we expect the r.v. to take? Such information can be provided by the *mean* and *standard deviation* of the r.v.
- First we consider discrete r.v.s
- Let $X \sim p_X(x)$. The expected value (or *mean*) of X is defined as

$$\mathrm{E}(X) = \sum_{x \in \mathcal{X}} x p_X(x)$$

Interpretations: If we view probabilities as relative frequencies, the mean would be the *weighted sum* of the relative frequencies.

If we view probabilities as point masses, the mean would be the *center of mass* of the set of mass points

- Example: If the weather is good, which happens with probability 0.6, Alice walks the 2 miles to class at a speed 5 miles/hr, otherwise she rides a bus at speed 30 miles/hr. What is the expected time to get to class?

  Solution: Define the discrete r.v. T to take the value $(2/5)$hr with probability 0.6 and $(2/30)$hr with probability 0.4. The expected value of T

$$\mathrm{E}(T) = 2/5 \times 0.6 + 2/30 \times 0.4 = 4/15 \text{ hr}$$

- The *second moment* (or *mean square* or *average power*) of X is defined as

$$\mathrm{E}(X^2) = \sum_{x \in \mathcal{X}} x^2 p_X(x) \geq 0 \quad \text{rms} = \sqrt{\mathrm{E}(X^2)}$$

For the previous example, the second moment is

$$\mathrm{E}(T^2) = (2/5)^2 \times 0.6 + (2/30)^2 \times 0.4 = 22/225 \text{ hr}^2$$

- The *variance* of X is defined as

$$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] = \sum_{x \in \mathcal{X}} (x - \mathrm{E}(X))^2 p_X(x) \geq 0$$

  Also often written σ_X^2. The variance has the interpretation of the *moment of inertia* about the center of mass for a set of mass points

- The *standard deviation* of X is defined as $\sigma_X = \sqrt{\mathrm{Var}(X)}$
- Variance in terms of mean and second moment: Expanding the square and using the linearity of sum, we obtain

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] = \sum_x (x - \mathrm{E}(X))^2 p_X(x) \\
&= \sum_x (x^2 - 2x\,\mathrm{E}(X) + [\mathrm{E}(X)]^2) p_X(x) \\
&= \sum_x x^2 p_X(x) - 2\,\mathrm{E}(X) \underbrace{\sum_x x p_X(x)}_{\mathrm{E}(X)} + [\mathrm{E}(X)]^2 \underbrace{\sum_x p_X(x)}_{1} \\
&= \mathrm{E}(X^2) - 2[\mathrm{E}(X)]^2 + [\mathrm{E}(X)]^2 \\
&= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 \geq 0
\end{aligned}
$$

Note that since for any r.v., $\mathrm{Var}(X) \geq 0$, $\mathrm{E}(X^2) \geq (\mathrm{E}(X))^2$

So, for our example, $\mathrm{Var}(T) = 22/225 - (4/15)^2 = 0.02667$.

Mean and Variance of Famous Discrete RVs

- Bernoulli r.v. $X \sim \mathrm{Bern}(p)$: The mean is

$$\mathrm{E}(X) = p \times 1 + (1 - p) \times 0 = p$$

The second moment is

$$\mathrm{E}(X^2) = p \times 1^2 + (1 - p) \times 0^2 = p$$

Thus the variance is

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

- Binomial r.v. $X \sim \mathrm{B}(n, p)$: It is not easy to find it using the definition. Later we use a much simpler method to show that

$$\mathrm{E}(X) = np$$
$$\mathrm{Var}(X) = np(1-p)$$

Just n times the mean (variance) of a Bernoulli!

- Geometric r.v. $X \sim \mathrm{Geom}(p)$:

One way to find the mean is to use standard calculus tricks for stuff. Do this for practice, but later find simpler methods using transforms.

If $|a| < 1$ then the geometric progression formula says that

$$\sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$$

Suppose instead you want to evaluate

$$\sum_{k=0}^{\infty} k a^k$$

How do you get the second sum from the first?

Just differentiate the first with respect to a and you almost get it:

$$\frac{d}{da} \sum_{k=0}^{\infty} a^k = \sum_{k=0}^{\infty} k a^{k-1} = \frac{1}{a} \sum_{k=0}^{\infty} k a^k$$

or

$$\sum_{k=0}^{\infty} ka^k = a\frac{d}{da}\frac{1}{1-a} = \frac{a}{(1-a)^2}$$

Thus

$$\begin{aligned}
E[X] &= \sum_{k=1}^{\infty} pk(1-p)^{k-1} \\
&= \frac{p}{1-p}\sum_{k=1}^{\infty} k(1-p)^k \\
&= \frac{p}{1-p}\sum_{k=0}^{\infty} k(1-p)^k \\
&= \frac{p}{1-p}\frac{1-p}{p^2} = \frac{1}{p}
\end{aligned}$$

The second moment and variance can be similarly found by taking two derivatives. Try it!

You should find that

$$\mathrm{E}(X^2) = \frac{2-p}{p^2}$$

and therefore

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{1-p}{p^2}$$

- Poisson r.v. $X \sim \text{Poisson}(\lambda)$: The mean is given by

$$
\begin{aligned}
\mathrm{E}(X) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} &= e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{(k-1)}}{(k-1)!} \\
&= \lambda e^{-\lambda}\underbrace{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}}_{e^{\lambda}} \\
&= \lambda
\end{aligned}
$$

Can show that the variance is also equal to λ

Mean and Variance for Continuous RVs

- Now consider a continuous r.v. $X \sim f_X(x)$, the expected value (or mean) of X is defined as

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

This has the interpretation of the center of mass for a mass density

- The *second moment* and variance are similarly defined as:

$$
\begin{aligned}
\mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \\
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}(X))^2\right] \\
&= \mathrm{E}(X^2) - (\mathrm{E}(X))^2
\end{aligned}
$$

- Uniform r.v. $X \sim \mathrm{U}[a,b]$: The mean, second moment, and variance are given by

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_a^b x \times \frac{1}{b-a}\,dx = \frac{a+b}{2} \\
\mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \int_a^b x^2 \times \frac{1}{b-a}\,dx \\
&= \frac{1}{b-a}\int_a^b x^2\,dx = \frac{1}{b-a} \times \frac{x^3}{3}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - (\mathrm{E}(X))^2 \\
&= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
\end{aligned}$$

Thus, for $X \sim \mathrm{U}[0,1]$, $\mathrm{E}(X) = 1/2$ and $\mathrm{Var} = 1/12$

- Exponential r.v. $X \sim \mathrm{Exp}(\lambda)$: The mean and variance are given by

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx \\
&= \int_0^{\infty} \underbrace{x}_{-u}\, \underbrace{\lambda\, \underbrace{e^{-\lambda x}}_{v}\,dx}_{-dv} \\
&= (-xe^{-\lambda x})\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\,dx \;\underbrace{\text{(integration by parts)}}_{\int udv = uv - \int vdu} \\
&= 0 - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty} = \frac{1}{\lambda} \\
\mathrm{E}(X^2) &= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,dx = \frac{2}{\lambda^2} \\
\mathrm{Var}(X) &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
\end{aligned}$$

- For a Gaussian r.v. $X \sim \mathcal{N}(\mu, \sigma^2)$, the mean is μ and the variance is σ^2 (will show this later using transforms, can prove using calculus, but it is messy)

Mean and Variance for Famous r.v.s

Random Variable	Mean	Variance
$\mathrm{Bern}(p)$	p	$p(1-p)$
$\mathrm{Geom}(p)$	$1/p$	$(1-p)/p^2$
$\mathrm{B}(n,p)$	np	$np(1-p)$
$\mathrm{Poisson}(\lambda)$	λ	λ
$\mathrm{U}[a,b]$	$(a+b)/2$	$(b-a)^2/12$
$\mathrm{Exp}(\lambda)$	$1/\lambda$	$1/\lambda^2$
$\mathcal{N}(\mu,\sigma^2)$	μ	σ^2

Cumulative Distribution Function (cdf)

- For discrete r.v.s we use pmf's, for continuous r.v.s we use pdf's
- Many real-world r.v.s are mixed, that is, have both discrete and continuous components

  Example: A packet arrives at a router in a communication network. If the input buffer is empty (happens with probability p), the packet is serviced immediately. Otherwise the packet must wait for a random amount of time as characterized by a pdf (e.g., Exp(λ))

  Define the r.v. X to be the packet service time. X is neither discrete nor continuous

- There is a third probability function that characterizes all random variable types — discrete, continuous, and mixed. The *cumulative distribution function* or *cdf* $F_X(x)$ of a random variable is defined by

$$F_X(x) = \mathrm{P}\{X \leq x\} \text{ for } x \in (-\infty, \infty)$$

- Properties of the cdf:
  - Like the pmf (but unlike the pdf), the cdf is the probability of something. Hence, $0 \leq F_X(x) \leq 1$
  - Normalization axiom implies that

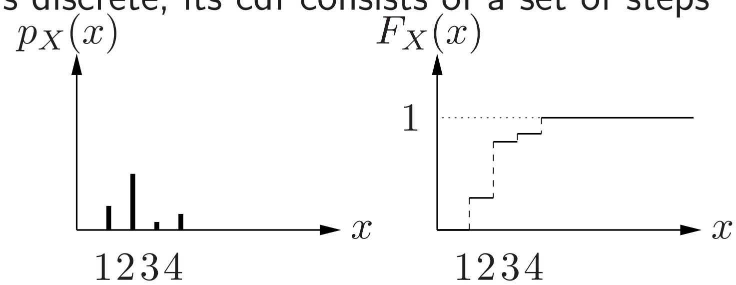
$$F_X(\infty) = 1, \text{ and } F_X(-\infty) = 0$$

- The probability of any event of the form $X \in (a, b]$ can be easily computed from a cdf, e.g.,

$$\begin{aligned} \mathrm{P}\{X \in (a, b]\} &= \mathrm{P}\{a < X \le b\} \\ &= \mathrm{P}\{X \le b\} - \mathrm{P}\{X \le a\} \text{ (additivity)} \\ &= F_X(b) - F_X(a) \end{aligned}$$

- Previous property $\Rightarrow$ $F_X(x)$ is monotonically nondecreasing, i.e., if $a > b$ then $F_X(a) \ge F_X(b)$
- The probability of any outcome a is: $\mathrm{P}\{X = a\} = \mathrm{P}\{X \le a\} - \mathrm{P}\{X < a\} = F_X(a) - F_X(a^-)$, where $F_X(a^-) = \lim_{x \uparrow a} F_X(a)$

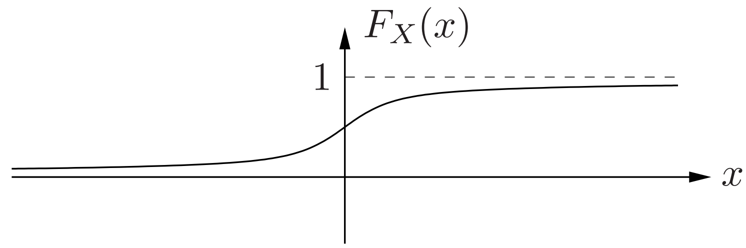
- If a r.v. is discrete, its cdf consists of a set of steps



- If X is a continuous r.v. with pdf $f_X(x)$, then

$$F_X(x) = \mathrm{P}\{X \le x\} = \int_{-\infty}^{x} f_X(\alpha)\, d\alpha$$

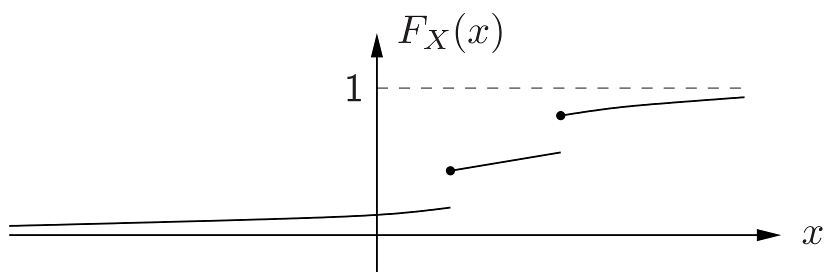
So, the cdf of a continuous r.v. X is continuous



In fact the precise way to define a continuous random variable is through the continuity of its cdf Further, if $F_X(x)$ is differentiable (almost everywhere), then

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{\mathrm{P}\{x < X \le x + \Delta x\}}{\Delta x}$$

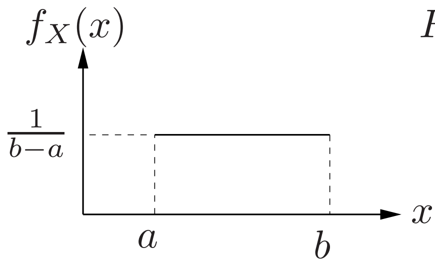
◦ The cdf of a mixed r.v. has the general form

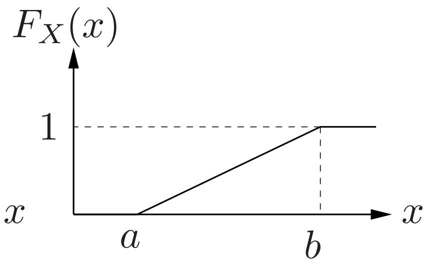


Examples

- cdf of a uniform r.v.:

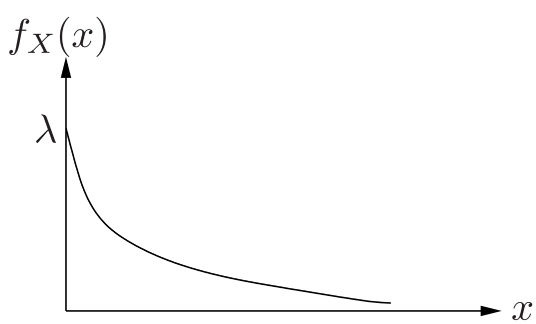
$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \int_a^x \frac{1}{b-a}\, d\alpha = \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x \ge b \end{cases}$$

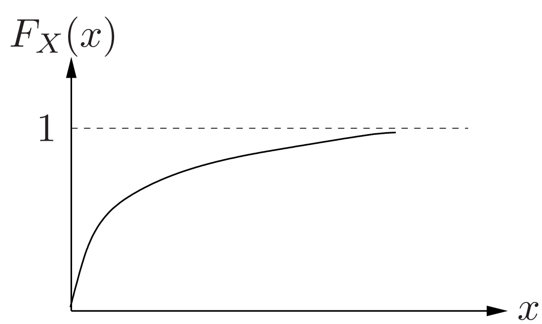




- cdf of an exponential r.v.:

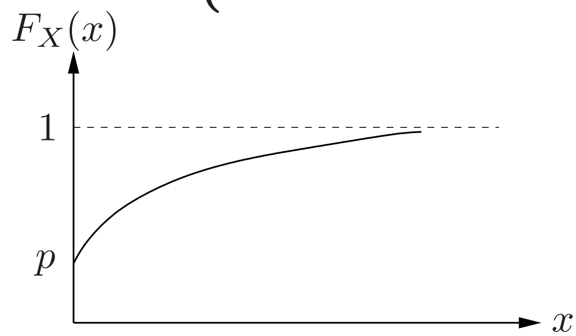
$$F_X(x) = 0,\ X < 0, \text{ and } F_X(x) = 1 - e^{-\lambda x},\ x \ge 0$$





- cdf for a mixed r.v.: Let X be the service time of a packet at a router. If the buffer is empty (happens with probability p), the packet is serviced immediately. If it is not empty, service time is described by an exponential pdf with parameter $\lambda > 0$

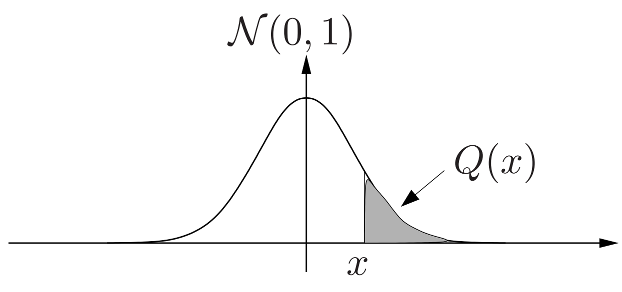
  The cdf of X is $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ p & \text{if } x = 0 \\ p + (1-p)\underbrace{(1 - e^{-\lambda x})}_{\text{cdf of Exp}(\lambda)} & \text{if } x > 0 \end{cases}$



- cdf of a Gaussian r.v.: There is no nice closed form for the cdf of a Gaussian r.v., but there are many published tables for the cdf of a standard normal pdf $\mathcal{N}(0,1)$, the Φ function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi$$

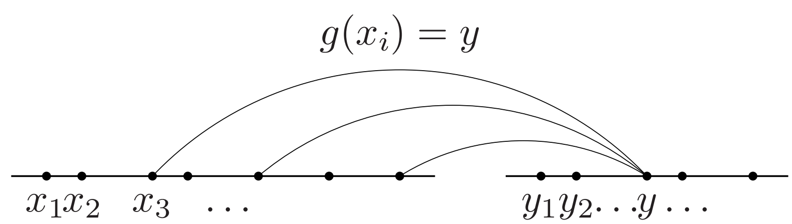
  More commonly, the tables are for the $Q(x) = 1 - \Phi(x)$ function



  Or, for the complementary error function: $\mathrm{erfc}(x) = 2Q(\sqrt{2}x)$ for $x > 0$ As we shall see, the $Q(\cdot)$ function can be used to quickly compute $\mathrm{P}\{X > a\}$ for any Gaussian r.v. X

Functions of a Random Variable

- We are often given a r.v. X with a known distribution (pmf, cdf, or pdf), and some function $y = g(x)$ of x, e.g., X^2, $|X|$, $\cos X$, etc., and wish to specify Y, i.e., find its pmf, if it is discrete, pdf, if continuous, or cdf
- Each of these functions is a random variable defined over the original experiment as $Y(\omega) = g(X(\omega))$. However, since we do not assume knowledge of the sample space or the probability measure, we need to specify Y directly from the pmf, pdf, or cdf of X
- First assume that $X \sim p_X(x)$, i.e., a discrete random variable, then Y is also discrete and can be described by a pmf $p_Y(y)$. To find it we find the probability of the inverse image $\{\omega : Y(\omega) = y\}$ for every y. Assuming Ω is discrete:



$$p_Y(y) = \mathrm{P}\{\omega : Y(\omega) = y\} = \sum_{\{\omega:\ g(X(\omega))=y\}} \mathrm{P}\{\omega\}$$

$$= \sum_{\{x:\ g(x)=y\}} \ \sum_{\{\omega:\ X(\omega)=x\}} \mathrm{P}\{\omega\} = \sum_{\{x:\ g(x)=y\}} p_X(x)$$

Thus

$$\boxed{p_Y(y) = \textstyle\sum_{\{x:\ g(x)=y\}} p_X(x)}$$

We can *derive* $p_Y(y)$ directly from $p_X(x)$ without going back to the original random experiment

- Example: Let the r.v. X be the maximum of two independent rolls of a four-sided die. Define a new random variable $Y = g(X)$, where

$$g(x) = \begin{cases} 1 & \text{if } x \geq 3 \\ 0 & \text{otherwise} \end{cases}$$

Find the pmf for Y

Solution: Recall $p_X : 1/16, 3/16, 5/16, 7/16$

$$p_Y(y) = \sum_{\{x:\, g(x)=y\}} p_X(x)$$

$$p_Y(1) = \sum_{\{x:\, x \geq 3\}} p_X(x)$$

$$= \frac{5}{16} + \frac{7}{16} = \frac{3}{4}$$

$$p_Y(0) = 1 - p_Y(1) = \frac{1}{4}$$

Derived Densities

- Let X be a continuous r.v. with pdf $f_X(x)$ and $Y = g(X)$ such that Y is a continuous r.v. We wish to find $f_Y(y)$
- Recall derived pmf approach: Given $p_X(x)$ and a function $Y = g(X)$, the pmf of Y is given by

$$p_Y(y) = \sum_{\{x:\, g(x)=y\}} p_X(x),$$

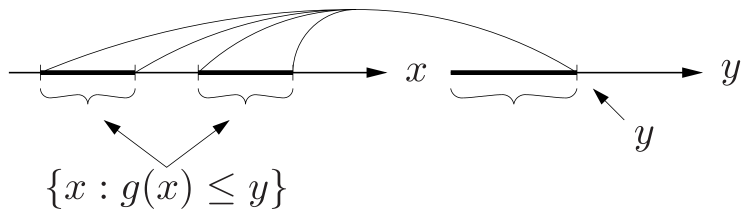
i.e., $p_Y(y)$ is the sum of $p_X(x)$ over all x that yield $g(x) = y$

- Idea does not extend immediately to deriving pdfs, since pdfs are not probabilities, and we cannot add probabilities of points

But basic idea does extend to cdfs

- Can first calculate the cdf of Y as

$$F_{\underbrace{Y}_{g(X)}}(y) = \mathrm{P}\{g(X) \leq y\} = \int_{\{x:\ g(x) \leq y\}} f_X(x)\, dx$$



- Then differentiate to obtain the pdf

$$f_Y(y) = \frac{dF_Y(y)}{dy}$$

Typically the hard part is getting the limits on the integral correct. Often they are obvious, but sometimes they are more subtle

Example: Squaring

- $X \sim f_X(x)$.
- $W = g(X) = X^2$, e.g., energy over a unit resistor of a voltage.
-

$$\begin{aligned} F_W(w) &= P(W \leq w) = P(X^2 \leq w) \\ &= P(-\sqrt{w} \leq X \leq \sqrt{w}) \\ &= \int_{-\sqrt{w}}^{\sqrt{w}} f_X(x)\, dx = F_X(\sqrt{w}) - F_X(-\sqrt{w}) \end{aligned}$$

Note the formulas assume that $w \geq 0$. If this is not true then the cdf is 0.

Rather than evaluate the integral for a specific input pdf, it is simpler to differentiate the cdf to directly find the pdf.

- The following differentiation formula is often useful for this purpose:

$$\frac{d}{dw}\int_{a(w)}^{b(w)} g(r)\,dr = g(b(w))\frac{db(w)}{dw} - g(a(w))\frac{da(w)}{dw}$$

Applying formula:

$$f_W(w) = \frac{f_X(\sqrt{w}) + f_X(-\sqrt{w})}{2\sqrt{w}};\ w > 0.$$

- Examples: $f_X(x)$ is uniform on $(0,1)$, then

$$f_W(w) = \frac{1}{2\sqrt{w}}\ ;\ w \in (0,1)$$

$f_X(x)$ is exponential with parameter λ, then

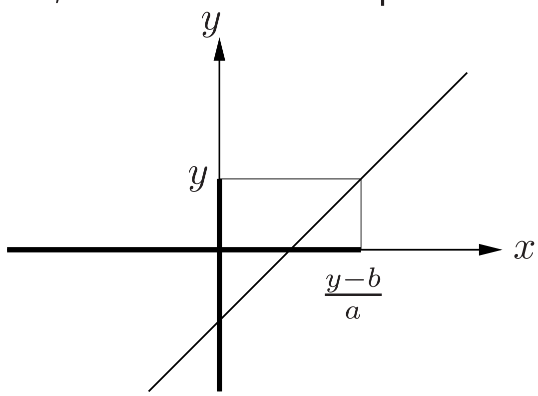
$$f_W(w) = \frac{\lambda e^{-\lambda\sqrt{w}}}{2\sqrt{w}}\ ;\ w > 0$$

$f_X(x)$ is $\mathcal{N}(0,\sigma_X^2)$, then

$$f_W(w) = \frac{e^{-w}}{\sqrt{2\pi\sigma_X w}}\ ;\ w > 0$$

Example: Linear Functions

- Let $X \sim f_X(x)$ and $Y = aX + b$ for some $a > 0$ and b. (The case $a < 0$ is left as an exercise)
- To find the pdf of Y, we use the above procedure



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

Thus

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

- Can show that for general $a \ne 0$,

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

• Example: $X \sim \mathrm{Exp}(\lambda)$, i.e.,

$$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0$$

Then

$$f_Y(y) = \begin{cases} \frac{\lambda}{|a|} e^{-\lambda(y-b)/a} & \text{if } \frac{y-b}{a} \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

• Example: $X \sim \mathcal{N}(\mu, \sigma^2)$, i.e.,

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

Again setting $Y = aX + b$,

$$\begin{aligned} f_Y(y) &= \frac{1}{|\, a \,|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|\, a \,|} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi(a\sigma)^2}} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} \quad \text{for } -\infty < y < \infty \end{aligned}$$

Therefore, $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

E.g., $X \sim \mathcal{N}(0,1)$, $Y = aX + b$, then $Y \sim \mathcal{N}(b, a^2)$

A linear (or affine) function of a Gaussian r.v. is another Gaussian r.v.!

This result can be used to compute probabilities for an arbitrary Gaussian r.v. from knowledge of the distribution a $\mathcal{N}(0,1)$ r.v.

Suppose $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and we wish to find $P\{a < Y \leq b\}$ for $b > a$ using tables. From the above result, we can express

$Y = \sigma_Y X + \mu_Y$, where $X \sim \mathcal{N}(0,1)$. Then

$$
\begin{aligned}
P\{a < Y \leq b\} &= P\{a < \sigma_Y X + \mu_Y \leq b\} \\
&= P\{a - \mu_Y < \sigma_Y X \leq b - \mu_Y\} \\
&= P\{\frac{a - \mu_Y}{\sigma_Y} < X \leq \frac{b - \mu_Y}{\sigma_Y}\} \\
&= P\{X \leq \frac{b - \mu_Y}{\sigma_Y}\} - P\{X \leq \frac{a - \mu_Y}{\sigma_Y}\} \\
&= \phi(\frac{b - \mu_Y}{\sigma_Y}) - \phi(\frac{a - \mu_Y}{\sigma_Y}) \\
&= Q(\frac{a - \mu_Y}{\sigma_Y}) - Q(\frac{b - \mu_Y}{\sigma_Y})
\end{aligned}
$$

Note Sometimes Q tables are only for positive arguments. Since the Gaussian density is symmetric, if $a > 0$ then $Q(a) = 1 - Q(-a)$ or $Q(-a) = 1 - Q(a)$. Thus if $(a - \mu_Y)/\sigma_Y < 0$,

$$P\{a < Y \leq b\} = 1 - Q(-\frac{a - \mu_Y}{\sigma_Y}) - Q(\frac{b - \mu_Y}{\sigma_Y})$$

The Q function

The $Q(\cdot)$ function is the area beneath the righthand tail of the Gaussian pdf $\mathcal{N}(0, 1)$:

$$Q(x) \;=\; \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \;.$$

The following table lists values of $Q(x)$ for $0 \leq x \leq 4$.

x	$Q(x)$	x	$Q(x)$
0.0	0.50000	2.0	2.2750×10^{-2}
0.1	0.46017	2.1	1.7864×10^{-2}
0.2	0.42074	2.2	1.3903×10^{-2}
0.3	0.38209	2.3	1.0724×10^{-2}
0.4	0.34458	2.4	8.1975×10^{-3}
0.5	0.30854	2.5	6.2097×10^{-3}
0.6	0.27425	2.6	4.6612×10^{-3}
0.7	0.24196	2.7	3.4670×10^{-3}
0.8	0.21186	2.8	2.5551×10^{-3}
0.9	0.18406	2.9	1.8658×10^{-3}
1.0	0.15866	3.0	1.3499×10^{-3}
1.1	0.13567	3.1	9.6760×10^{-4}
1.2	0.11507	3.2	6.8714×10^{-4}
1.3	0.09680	3.3	4.8342×10^{-4}
1.4	0.08076	3.4	3.3693×10^{-4}
1.5	0.06681	3.5	2.3263×10^{-4}
1.6	0.05480	3.6	1.5911×10^{-4}
1.7	0.04457	3.7	1.0780×10^{-4}
1.8	0.03593	3.8	7.2348×10^{-5}
1.9	0.02872	3.9	4.8096×10^{-5}
2.0	0.02275	4.0	3.1671×10^{-5}

Example: A Nonlinear Function

- John is driving a distance of 180 miles at constant speed that is uniformly distributed between 30 and 60 miles/hr. What is the pdf of the duration of the trip?
- Solution: Let X be John's speed, then

$$f_X(x) = \begin{cases} 1/30 & \text{if } 30 \le x \le 60 \\ 0 & \text{otherwise,} \end{cases}$$

The duration of the trip is $Y = 180/X$

To find $f_Y(y)$ we first find $F_Y(y)$. Note that $\{y : Y \leq y\} = \{x : X \geq 180/y\}$, thus

$$F_Y(y) = \int_{180/y}^{\infty} f_X(x)\, dx$$
$$= \begin{cases} 0 & \text{if } y \leq 3 \\ \int_{180/y}^{60} f_X(x)\, dx & \text{if } 3 < y \leq 6 \\ 1 & \text{if } y > 6 \end{cases}$$
$$= \begin{cases} 0 & \text{if } y \leq 3 \\ (2 - \frac{6}{y}) & \text{if } 3 < y \leq 6 \\ 1 & \text{if } y > 6 \end{cases}$$

Differentiating, we obtain

$$f_Y(y) = \begin{cases} 6/y^2 & \text{if } 3 \leq y \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Monotonic Functions

- Let $g(x)$ be a monotonically increasing and differentiable function over its range

  Then g is invertible, i.e., there exists a function h, such that

  $$y = g(x) \text{ if and only if } x = h(y)$$

  Often this is written as g^{-1}. If a function g has these properties, then $g(x) \leq y$ iff $x \leq h(y)$, and we can write

  $$\begin{aligned} F_Y(y) &= \mathrm{P}\{g(X) \leq y\} \\ &= \mathrm{P}\{X \leq h(y)\} \\ &= F_X(h(y)) \\ &= \int_{-\infty}^{h(y)} f_X(x)\, dx \end{aligned}$$

Thus
$$f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(h(y))\, \frac{dh}{dy}\ (y)$$

- Generalizing the result to both monotonically increasing and decreasing functions yields

  $$f_Y(y) = f_X(h(y)) \left| \frac{dh}{dy}\ (y) \right|$$

- Example: Recall the $X \sim \mathrm{U}[30, 60]$ example with $Y = g(X) = 180/X$

  The inverse is $X = h(Y) = 180/Y$

Applying the above formula in region of interest $Y \in [3,6]$ (it is 0 outside) yields

$$
\begin{aligned}
f_Y(y) &= f_X(h(y)) \left| \frac{dh}{dy}(y) \right| \\
&= \frac{1}{30}\frac{180}{y^2} \\
&= \begin{cases} \frac{6}{y^2} & \text{for } 3 \le y \le 6 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- Another Example:

  Suppose that X has a pdf that is nonzero only in $[0,1]$ and define $Y = g(X) = X^2$

  In the region of interest the function is invertible and $X = \sqrt{Y}$

  Applying the pdf formula, we obtain

$$
\begin{aligned}
f_Y(y) &= f_X(h(y)) \left| \frac{dh}{dy}(y) \right| \\
&= \frac{f_X(\sqrt{y})}{2\sqrt{y}}, \text{ for } 0 < y \le 1
\end{aligned}
$$

- Personally I prefer the more fundamental approach, since I often forget this formula or mess up the signs