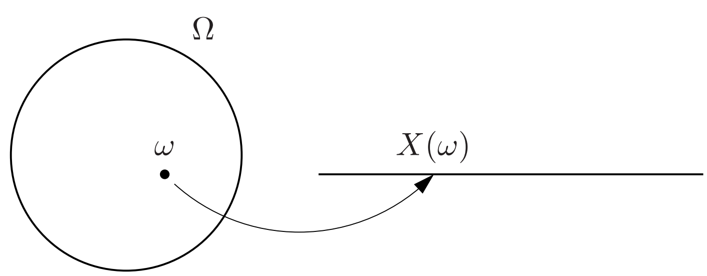
Lecture Notes 2
Random Variables

- Definition
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)
- Probability Density Function (PDF)
- Functions of a Random Variable
- Application: Generation of Random Variables

Random Variable

- A *random variable* (r.v.) is a real-valued function $X(\omega)$ over a sample space Ω, i.e., $X : \Omega \to \mathbf{R}$



- Notations:
  - We use upper case letters for random variables: X, Y, Z, Φ, Θ, . . .
  - We use lower case letters for *values* of random variables: $X = x$ means that random variable X takes on the value x, i.e., $X(\omega) = x$ where ω is the outcome

Examples of Random Variables

1. Let the random variable X be the number of heads in n coin flips. The sample space is $\Omega = \{H, T\}^n$, the possible outcomes of n coin flips; then

$$X \in \{0, 1, 2, \ldots, n\}$$

2. Let $\Omega = \mathbf{R}$, the real numbers. Define random variables X and Y as follows:

   a. $X(\omega) = \omega$

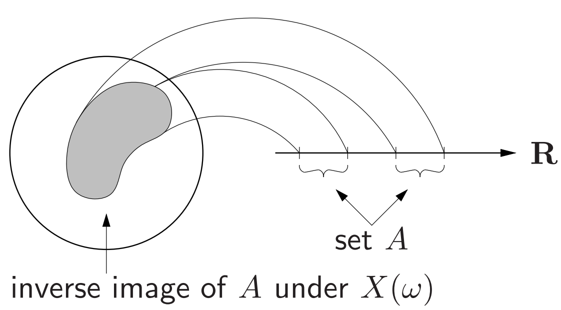
   b. $Y(\omega) = \begin{cases} +1 & \omega \geq 0 \\ -1 & \text{otherwise} \end{cases}$

3. Packet arrival times in the interval $(0, T\,]$. Here Ω is the set of all finite length strings $(t_1, t_2, \ldots, t_n) \in (0, T\,]^*$, where $t_1 \leq t_2 \leq \cdots \leq t_n$. Define the random variable X to be n, the length of the string; then $X \in \{0, 1, 2, 3, \ldots\}$

4. Let X be the service time at a router. If the buffer is empty the packet is served immediately, i.e., $X = 0$. If it is not empty, the service time $X > 0$ is a positive real number

Specifying a Random Variable

- Specifying a random variable means being able to determine the probability that $X \in A$ for any Borel set $A \subset \mathbf{R}$, in particular, for any interval $(a, b\,]$
- To do so, consider the *inverse image* of A under X, i.e., $\{\omega : X(\omega) \in A\}$



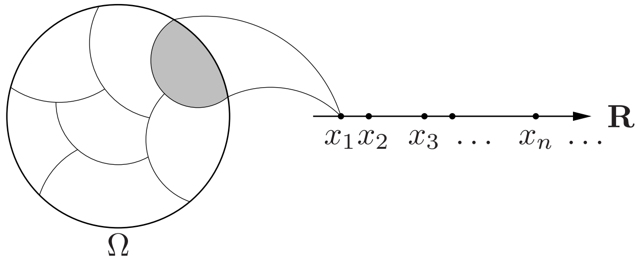
- Since $X \in A$ iff $\omega \in \{\omega : X(\omega) \in A\}$,

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

  Shorthand: $\mathrm{P}(\{\text{set description}\}) = \mathrm{P}\{\text{set description}\}$

Discrete Random Variables

- A random variable is said to be *discrete* if $\mathrm{P}\{X \in \mathcal{X}\} = 1$ for some *countable* set $\mathcal{X} \subset \mathbf{R}$, i.e., $\mathcal{X} = \{x_1, x_2, \ldots\}$ (finite or infinite)
- Examples 1, 2b, and 3 on page 2-3 are discrete random variables
- In general, $X(\omega)$ partitions Ω into the sets $\{\omega : X(\omega) = x_i\}$, for $i = 1, 2, \ldots$



In order to specify X, it suffices to know $\mathrm{P}\{X = x_i\}$ for every i

- A discrete random variable is thus completely specified by its *probability mass function* (pmf)
$$p_X(x) \;=\; \mathrm{P}\{X = x\} \text{ for every } x \in \mathcal{X}$$
- Clearly $p_X(x) \geq 0$ and $\sum_{x \in \mathcal{X}} p_X(x) = 1$
- Note that $p_X(x)$ can be simply viewed as a probability measure over a discrete sample space (even though the original sample space may be continuous as in examples 2b and 3)
- The probability of any (Borel) set $A \subset \mathbf{R}$ is given by
$$\mathrm{P}\{X \in A\} = \sum_{x \in A \cap \mathcal{X}} p_X(x)$$
- 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 $p(x)$

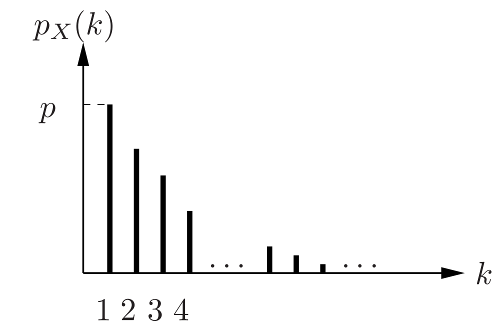
Famous Discrete Random Variables

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

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

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

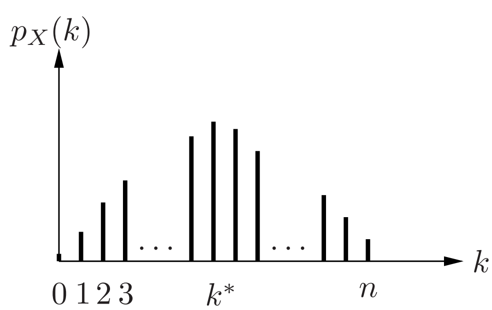
$$p_X(k) = p(1-p)^{k-1}\,, \quad k = 1, 2, 3, \ldots$$



The geometric r.v. represents, for example, the number of coin flips until the first heads occurs (assuming independent coin flips)

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

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



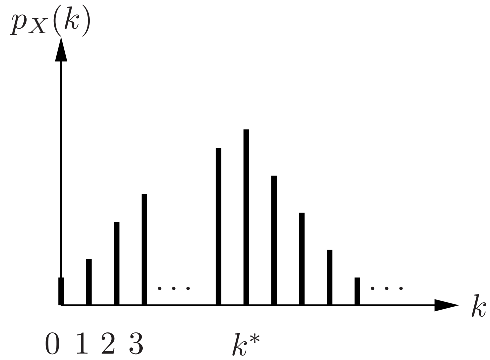
The binomial r.v. represents, for example, the number of heads that occur in n independent coin flips

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

$$k^* = \begin{cases} (n+1)p \ \text{ and } \ (n+1)p - 1 & \text{if } (n+1)p \text{ is an integer} \\ \lfloor (n+1)p \rfloor & \text{otherwise} \end{cases}$$

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

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



The Poisson r.v. represents the number of random events in a unit time, e.g., arrivals of packets, photons, customers — λ is the average arrival rate

The maximum of $p_X(k)$ attained at

$$k^* = \begin{cases} \lambda \text{ and } \lambda - 1 & \text{if } \lambda \text{ is an integer} \\ \lfloor \lambda \rfloor & \text{otherwise} \end{cases}$$

- Fact: Poisson is the limit of Binomial when $p \propto \frac{1}{n}$ as $n \to \infty$

To show this, let $X_n \sim \text{Binom}(n, \lambda/n)$ where $\lambda > 0$

For any fixed nonnegative integer k,

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

Cumulative Distribution Function

- To specify a random variable, we need to be able to determine $\mathrm{P}\{X \in A\}$ for any Borel set $A \subset \mathbf{R}$, i.e., any set generated by starting from intervals and taking countable unions, intersections, and complements
- It suffices to specify $\mathrm{P}\{X \in (a, b]\}$ for all intervals. The probability of any other Borel set can be determined by the axioms of probability
- Equivalently, it suffices to specify its *cumulative distribution function* (cdf):

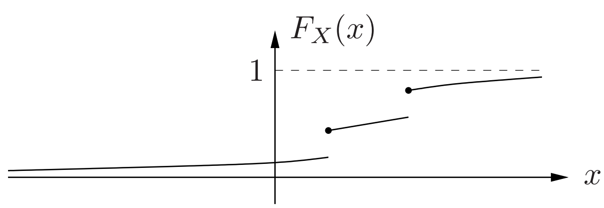
$$F_X(x) = \mathrm{P}\{X \le x\} = \mathrm{P}\{X \in (-\infty, x]\}, \quad x \in \mathbf{R}$$

To see the equivalence: if $a < b$ then

$$\mathrm{P}\{X \in (a, b]\} = \mathrm{P}\{X \le b\} - \mathrm{P}\{X \le a\} = F_X(b) - F_X(a)$$

Properties of CDF

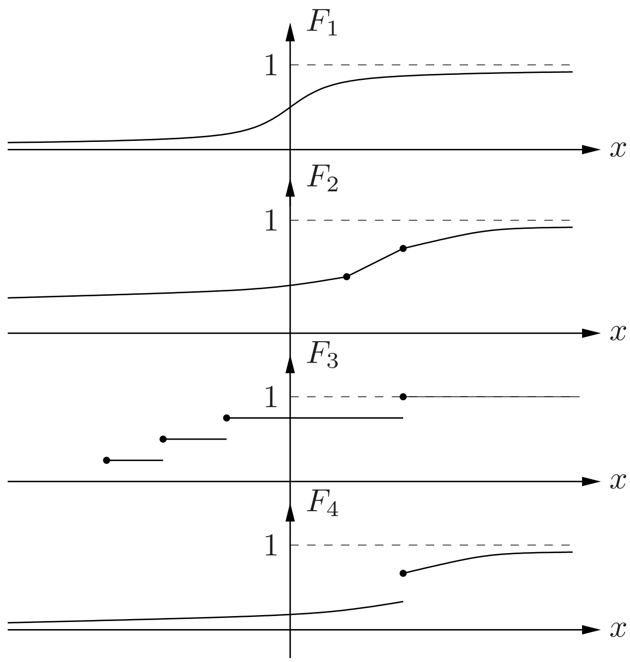
- $F_X(x) \ge 0$
- $F_X(x)$ is monotonically nondecreasing, i.e., if $a > b$ then $F_X(a) \ge F_X(b)$
- Limits:
  - $\lim_{x \to +\infty} F_X(x) = 1$
  - $\lim_{x \to -\infty} F_X(x) = 0$



- $F_X(x)$ is right continuous, i.e., $F_X(a^+) = \lim_{x \to a^+} F_X(x) = F_X(a)$
- $\mathrm{P}\{X = a\} = F_X(a) - F_X(a^-)$, where $F_X(a^-) = \lim_{x \to a-} F_X(x)$

- For any Borel set A, $\mathrm{P}\{X \in A\}$ can be determined from $F_X(x)$
- For a discrete random variable, $F_X(x)$ consists only of a countable set of steps
- A random variable is said to be *continuous* if its cdf is a continuous function (e.g., Example 2a on page 2-3)
- A random variable is said to be *mixed* if it is neither discrete nor continuous (e.g., Example 4 on page 2-3)

Example CDFs



Probability Density Function

- If $F_X(x)$ is continuous and differentiable (except possibly over a countable set), then X has a *probability density function* (pdf) $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$$

- If $F_X(x)$ is differentiable everywhere, then (by definition of derivative)

$$\begin{aligned} f_X(x) &= \frac{dF_X(x)}{dx} \\ &= \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\mathrm{P}\{x < X \le x + \Delta x\}}{\Delta x} \end{aligned}$$

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

Properties of PDF

- $f_X(x) \ge 0$
- $\displaystyle\int_{-\infty}^{\infty} f_X(x)\, dx = 1$
- For any event (Borel set) $A \subset \mathbf{R}$,

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

  In particular,

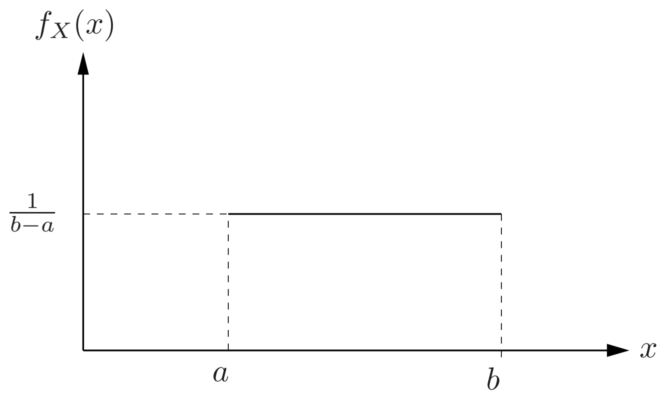
$$\mathrm{P}\{x_1 < X \le x_2\} = \int_{x_1}^{x_2} f_X(x)\, dx$$

- Important note: $f_X(x)$ should not be interpreted as the probability that $X = x$. In fact, $f_X(x)$ is *not* a probability measure since it can be > 1
- Remark: We can use delta functions to define a pdf for a discrete or a mixed random variable, but this is not commonly done in the field of probability

Famous Continuous Random Variables

- *Uniform*: $X \sim \mathrm{U}[a,b]$ where $a < b$ has pdf

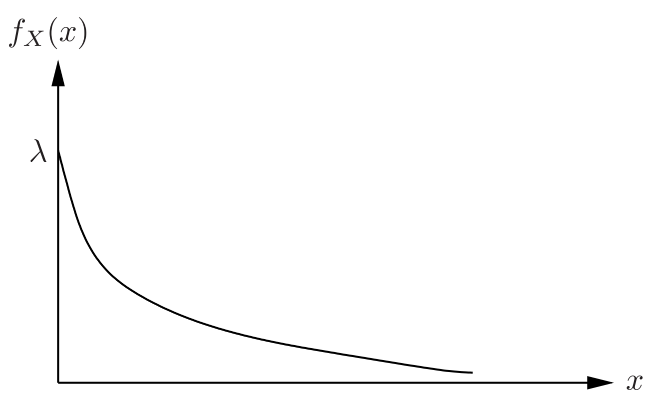
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



The uniform r.v. is commonly used in modeling quantization noise and finite precision computation error (roundoff error)

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

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



The exponential r.v. represents *interarrival time* in a queue (time between two consecutive packet or customer arrivals) or service time in a queue, or particle lifetime

- Example: Let $X \sim \text{Exp}(0.1)$ be the customer service time in minutes at a bank

  The person ahead of you has been served for 10 minutes. What is the probability that you will wait another 10 minutes or more before getting served?

  Solution: We want to find $\text{P}\{X > 20 \,|\, X > 10\}$

  By definition,

$$\begin{aligned} \text{P}\{X > 20 \,|\, X > 10\} &= \frac{\text{P}\{X > 20, X > 10\}}{\text{P}\{X > 10\}} \\ &= \frac{\text{P}\{X > 20\}}{\text{P}\{X > 10\}} = \frac{e^{-2}}{e^{-1}} = e^{-1} \end{aligned}$$

  But $\text{P}\{X > 10\} = e^{-1}$. Therefore the conditional probability of waiting more than 10 minutes is the same as the *un*conditional probability of waiting more than 10 minutes!

- In general for any exponential r.v., whenever $0 \le x' < x$,

$$\text{P}\{X > x \,|\, X > x'\} = \text{P}\{X > x - x'\}$$

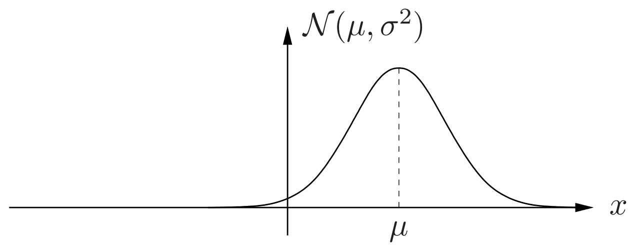
  Because of this property, the exponential r.v. is called *memoryless*

Gaussian Random Variable

- *Gaussian*: $X \sim \mathcal{N}(\mu, \sigma^2)$ with parameters μ (the *mean*) and σ^2 (the *variance*) (or the *standard deviation* σ) has pdf

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

  The graph of the Gaussian (or *normal*) density is the bell-shaped curve



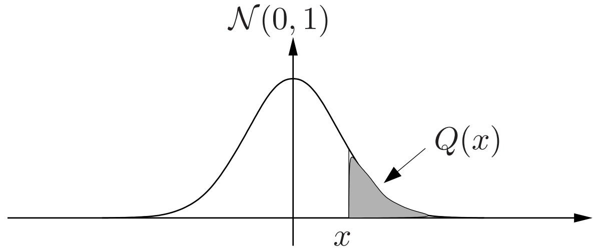
- The Gaussian r.v. is frequently encountered in nature — thermal and shot noise in electronic devices are Gaussian — and very frequently used in modelling various social, biological, and other phenomena. A *lot* more on Gaussian r.v.s later

Φ, Q, and erfc Functions

- The cdf of the standard normal random variable $\mathcal{N}(0,1)$ is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

- Also define the function $Q(x) = 1 - \Phi(x) = \mathrm{P}\{X > x\}$



The $Q(\cdot)$ function can be used to compute $\mathrm{P}\{X > a\}$ for *any* Gaussian r.v. X

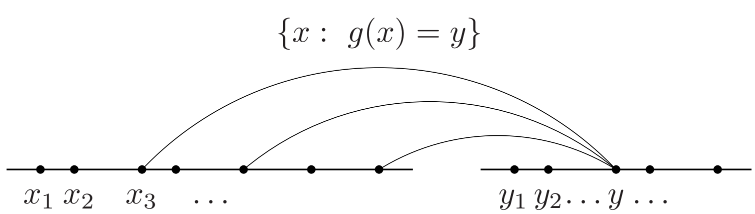
- The *complementary error function* is

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-u^2} du = 2Q(\sqrt{2}\, x)$$

Functions of a Random Variable

- Suppose we are given a r.v. with known distribution (pmf, cdf, or pdf), a function $y = g(x)$, and want to specify the random variable $Y = g(X)$
- If $X \sim p_X(x)$ is a discrete r.v., then Y is also discrete with pmf

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



Derived Densities

- If $X \sim f_X(x)$ is a continuous r.v., and $g(x)$ is differentiable, then we can find the pdf of Y by either of two methods, using the approximations of probabilities by densities times differentials, or directly by finding the cdf and then differentiating.

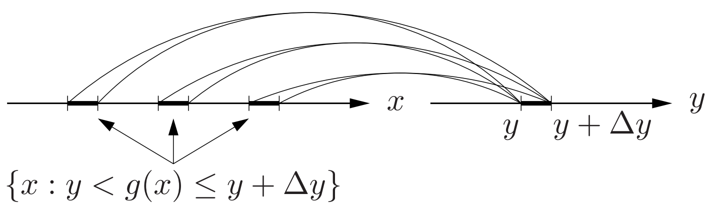
  First method: By definition of pdf, $f_Y(y) = \lim_{\Delta y \to 0} \frac{\mathrm{P}\{y < Y \leq y + \Delta y\}}{\Delta y}$

  To find f_Y, we find

  $$\mathrm{P}\{y < Y \leq y + \Delta y\} = \mathrm{P}\{x : y < g(x) \leq y + \Delta y\},$$

  the probability of the inverse image under $g(x)$ of $(y, y + \Delta y\,]$, then take limits

  $$f_Y(y) \;=\; \lim_{\Delta y \to 0} \frac{\mathrm{P}\{x : y < g(x) \leq y + \Delta y\}}{\Delta y}$$



Second method: First find the cdf

$$F_Y(y) = \mathrm{P}\{Y \leq y\} = \mathrm{P}\{g(X) \leq y\} = \int_{x : g(x) \leq y} f_X(x)dx$$

and then find the pdf by differentiation

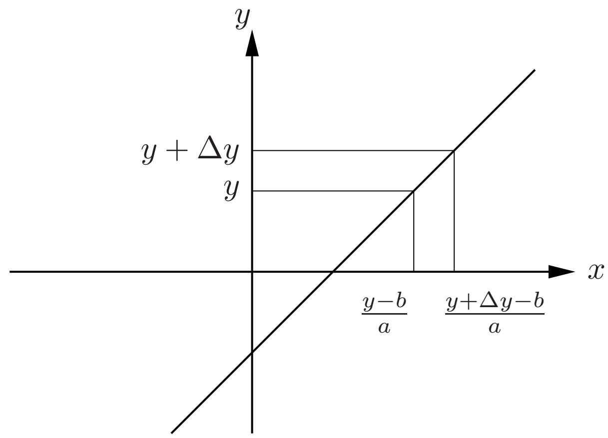
$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

The differentiation is usually simple as y will appear only in the limits of the integral and not in the integrand, and one can use the calculus formula for differentiating such an integral:

$$\frac{d}{dy} \int_{a(y)}^{b(y)} h(r)dr = h(b(y))\frac{db(y)}{dy} - h(a(y))\frac{da(y)}{dy}$$

We find $f_Y(y)$ for two examples, then give a general formula for $f_Y(y)$

- Example: *Linear function*. Let $X \sim f_X(x)$ and $Y = aX + b,\ a > 0$. Find $f_Y(y)$
  First method:



$$f_Y(y)\Delta y \approx \mathrm{P}\{x : y < g(x) \le y + \Delta y\}$$
$$= \mathrm{P}\left\{\frac{y-b}{a} < X \le \frac{y-b}{a} + \frac{\Delta y}{a}\right\} \approx f_X\left(\frac{y-b}{a}\right)\frac{\Delta y}{|a|}$$

As we let $\Delta y \to 0$, we obtain $f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$

Second method: If $a > 0$

$$\begin{aligned} F_Y(y) &= \int_{x:ax+b\le y} f_X(x)dx \\ &= \int_{x:x\le (y-b)/a} f_X(x)dx \\ f_Y(y) &= \frac{d}{dy}\int_{-\infty}^{(y-b)/a} f_X(x)dx \\ &= \frac{f_X(\frac{y-b}{a})}{a} \end{aligned}$$

Modifying the argument for negative a yields the same result as before, without approximations or limits.

- Special case: $X \sim \mathcal{N}(\mu, \sigma^2)$, i.e.,

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

Again setting $Y = aX + b$, for $-\infty < y < \infty$,

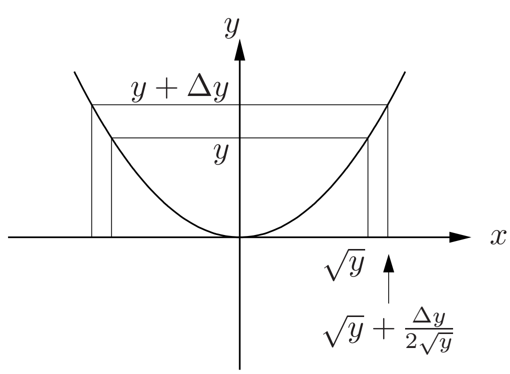
$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi(a\sigma)^2}}\, e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}} \end{aligned}$$

Therefore, $Y \sim \mathcal{N}(a\mu + b,\, a^2\sigma^2)$, i.e., $\sigma_Y^2 = a^2\sigma^2$ and $\mu_Y = a\mu + b$

This result can be used to compute probabilities for an arbitrary Gaussian r.v. from the distribution of the $\mathcal{N}(0,1)$ r.v. (i.e., using the $Q(\cdot)$ function)

- Example: *Quadratic function.* Let $X \sim f_X(x)$ and $Y = X^2$. Find $f_Y(y)$

First method:



$$\begin{aligned} f_Y(y)\Delta y &\approx \mathrm{P}\{x : y < g(x) \le y + \Delta y\} \\ &= \mathrm{P}\left\{+\sqrt{y} < X \le +\sqrt{y} + \frac{\Delta y}{2\sqrt{y}} \quad \text{or} \ -\sqrt{y} - \frac{\Delta y}{2\sqrt{y}} < X \le -\sqrt{y}\right\} \\ &\approx \left(\frac{1}{2\sqrt{y}} f_X(+\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})\right)\Delta y \end{aligned}$$

Therefore

$$f_Y(y) = \frac{1}{2\sqrt{y}}\Big(f_X(+\sqrt{y}) + f_X(-\sqrt{y})\Big)$$

Second method:

$$\begin{aligned} F(Y(y) &= P\{X^2 \le y\} = \int_{x:x^2\le y} f_X(x)dx \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \\ f_Y(y) &= \frac{d}{dy}\int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \\ &= f_X(\sqrt{y})\left(\frac{y^{-1/2}}{2}\right) - f_X(-\sqrt{y})\left(\frac{-y^{-1/2}}{2}\right) \\ &= \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right) \end{aligned}$$

as before.

In the special case where X is Gaussian with zero mean and variance σ^2, $f_Y(y) = e^{-y/2\sigma^2}/\sqrt{2\pi\sigma^2 y}$ for $y \ge 0$, the *chi-squared pdf with one degree of*

freedom.

- In general, let $X \sim f_X(x)$ and $Y = g(X)$ be differentiable. Then

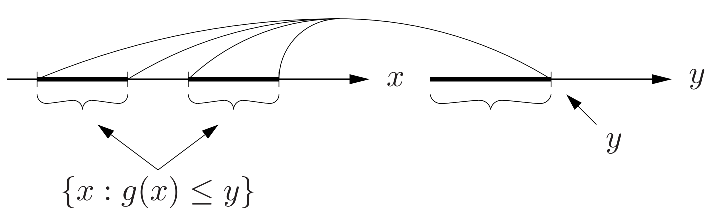
$$f_Y(y) = \sum_{i=1}^{k} \frac{f_X(x_i)}{|g'(x_i)|},$$

where $x_1, x_2, \ldots$ are the solutions of the equation $y = g(x)$ and $g'(x_i)$ is the derivative of g evaluated at x_i

Derived CDFs

- If the cdf of X is given and we wish to find the cdf of Y, then we use

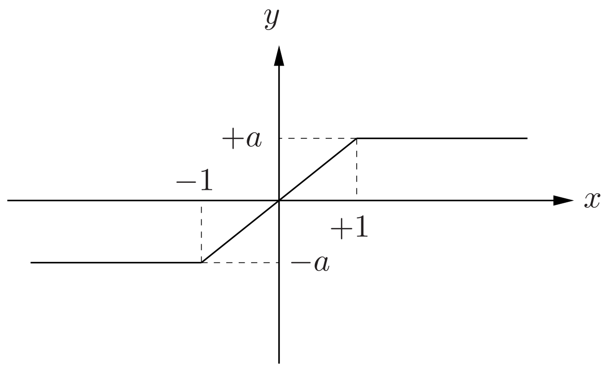
$$F_Y(y) = \mathrm{P}\{Y \leq y\} = \mathrm{P}\{x : g(x) \leq y\}$$



- This method is needed in cases where X does not have a density or the function $g(x)$ is not differentiable (or both)

  In some cases it is also easier to use when X has a density and $g(x)$ is not differentiable (in this case we find $F_Y(y)$ first then take derivatives to find $f_Y(y)$)

• Example: *Limiter*. Let X be a r.v. with Laplacian pdf $f_X(x) = \frac{1}{2}e^{-|x|}$, and let Y be defined by the function of X shown in the figure. Find the cdf of Y



Solution: To find the cdf of Y, we consider the following cases

◦ $y < -a$: Here clearly $F_Y(y) = 0$

◦ $y = -a$: Here

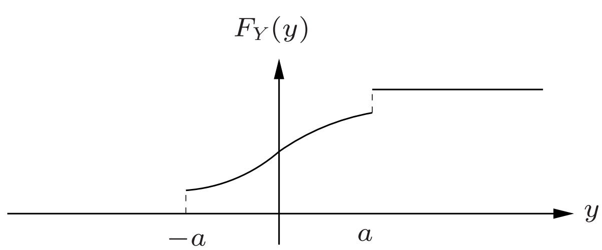
$$\begin{aligned} F_Y(-a) &= F_X(-1) \\ &= \int_{-\infty}^{-1} \tfrac{1}{2}e^{x}\,dx = \tfrac{1}{2}e^{-1} \end{aligned}$$

◦ $-a < y < a$: Here

$$\begin{aligned} F_Y(y) &= \mathrm{P}\{Y \le y\} \\ &= \mathrm{P}\left\{aX \le y\right\} \\ &= \mathrm{P}\left\{X \le \frac{y}{a}\right\} = F_X\left(\frac{y}{a}\right) \\ &= \tfrac{1}{2}e^{-1} + \int_{-1}^{y/a} \tfrac{1}{2}e^{-|x|}\,dx \end{aligned}$$

◦ $y \ge a$: Here $F_Y(y) = 1$

Combining the results, the following is a sketch of the cdf of Y



Application: Generation of Random Variables

- Generating a r.v. with a prescribed distribution is often needed for performing simulations involving random phenomena, e.g., noise or random arrivals
- First let $X \sim F(x)$ where the cdf $F(x)$ is continuous and strictly increasing. Define $Y = F(X)$, a real-valued random variable that is a function of X

  What is the cdf of Y?

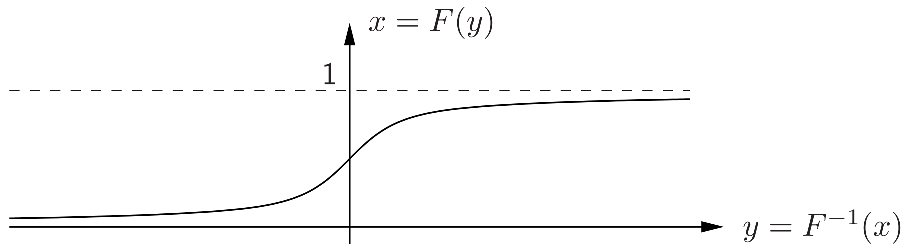
  Clearly, $F_Y(y) = 0$ for $y < 0$, and $F_Y(y) = 1$ for $y > 1$

  For $0 \le y \le 1$, note that by assumption F has an inverse F^{-1}, so

  $$F_Y(y) = \mathrm{P}\{Y \le y\} = \mathrm{P}\{F(X) \le y\} = \mathrm{P}\{X \le F^{-1}(y)\} = F(F^{-1}(y)) = y$$

  Thus $Y \sim U[\,0,1\,]$, i.e., Y is a uniformly distributed random variable
- Note: $F(x)$ does not need to be invertible. If $F(x) = a$ is constant over some interval, then the probability that X lies in this interval is zero. Without loss of generality, we can take $F^{-1}(a)$ to be the leftmost point of the interval
- Conclusion: We can generate a $\mathrm{U}[\,0,1\,]$ r.v. from *any* continuous r.v.

- Now, let's consider the more useful scenario where we are given $X \sim \mathrm{U}[\,0,1\,]$ (a random number generator) and wish to generate a random variable Y with prescribed cdf $F(y)$, e.g., Gausssian or exponential



- If F is continuous and strictly increasing, set $Y = F^{-1}(X)$. To show $Y \sim F(y)$,

  $$\begin{aligned} F_Y(y) &= \mathrm{P}\{Y \le y\} \\ &= \mathrm{P}\{F^{-1}(X) \le y\} \\ &= \mathrm{P}\{X \le F(y)\} \\ &= F(y)\,, \end{aligned}$$

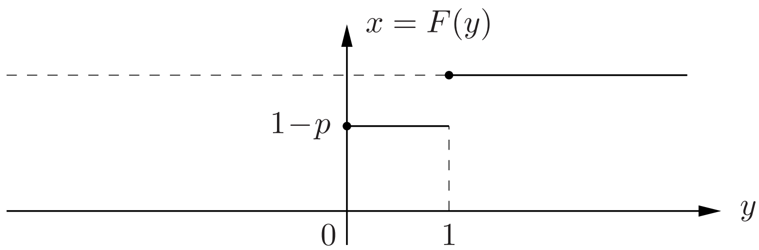
  since $X \sim \mathrm{U}[\,0,1\,]$ and $0 \le F(y) \le 1$

- Example: To generate $Y \sim \mathrm{Exp}(\lambda)$, set

$$Y = -\frac{1}{\lambda}\ln(1-X)$$

- Note: F does not need to be continuous for the above to work. For example, to generate $Y \sim \mathrm{Bern}(p)$, we set

$$Y = \begin{cases} 0 & X \le 1-p \\ 1 & \text{otherwise} \end{cases}$$



- Conclusion: We can generate a r.v. with *any* desired distribution from a $\mathrm{U}[0,1]$ r.v.