Essential Mathematical Methods for Engineers

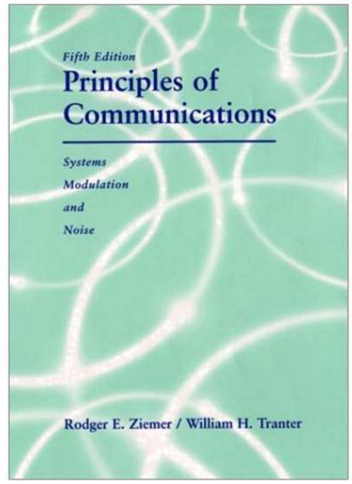
Lecture 5:
Probability and random variables

Outline

- sample spaces and the axioms of probability
- random variables and related functions
  - random variables
  - probability (cumulative) distribution functions
  - probability density functions
  - joint cdfs and pdfs
  - transformations of random variables
- statistical averages
  - average of a discrete random variable
  - average of a continuous random variable
  - average of a function of a random variable
  - average of a function of multiple random variables
  - variance of a random variable
  - average of a linear combination of N random variables
  - variance of a linear combination of independent random variables
  - the characteristic function
  - the PDF of the sum of two independent random variables
  - covariance and correlation coefficients

Outline

- some useful PDFs
  - binomial distribution
  - Poisson distribution
  - geometric distribution
  - Gaussian distribution
  - Rayleigh distribution
  - exponential distribution

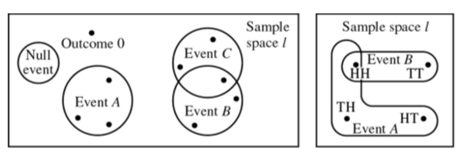


The material in this lecture is adapted from Principles of Communications, 5th ed., Ziemer & Tranter, Wiley

Sample spaces and the axioms of probability

- a chance experiment viewed geometrically
  - all possible outcomes define the sample space, S
  - an event is a collection of outcomes
  - an impossible collection of outcomes is defined as the null event

- some useful notations
  - $A \cup B$ is the event A or B or both
  - $A \cap B$ is the event A and B
  - $\overline{A}$ is the event "not A"



- a set of satisfactory axioms is the following
  - Axiom 1. $P(A) \geq 0$ for all events A in the sample space S
  - Axiom 2. The probability of all possible events occurring, $\mathrm{P}(S) = 1$.
  - Axiom 3. If the occurrence of A precludes the occurrence of B, and vice versa (i.e., A and B are mutually exclusive), then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$

Random variables

- continuous random variables
  - e.g. noise
- discrete random variables
  - e.g. the tossing of a coin
- notation
  - capital letters denote a random variable
    - e.g. X, Θ and so on
  - lowercase letters denote the value that the random variable takes on
    - e.g. x, θ and so on

Probability (cumulative) distribution functions

- a probabilistic description of random variables
- the cdf $F_X(x)$ is defined as

  $F_X(x)$ = probability that $X \leq x = P(X \leq x)$
  - $F_X(x)$ is a function of x, not of the random variable X
- properties:
  - $0 \leq F_X(x) \leq 1$, with $F_X(-\infty) = 0$ and $F_X(\infty) = 1$.
  - $F_X(x)$ is continuous from the right, that is, $\lim_{X \to X_{0^+}} F_X(x) = F_X(x_0)$.
  - $F_X(x)$ is a nondecreasing function of x; that is $F_X(x_1) \leq F_X(x_2)$ if $x_1 < x_2$

Random variables and related functions

Probability density functions

- for the purposes of computing averages the pdf, $f_X(x)$ is more useful
  - it is defined as $$f_X(x) = \frac{dF_X(x)}{dx}$$
  - thus $$F_X(x) = \int_{-\infty}^{x} f_X(\eta)d\eta$$
- the pdf has the following properties

$$f_X(x) = \frac{dF_X(x)}{dx} \geq 0 \qquad \int_{-\infty}^{\infty} f_X(x)dx = 1$$

$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx$$

- and by setting $x_1 = x - dx$ and $x_2 = x$

$$f_X(x)dx = P(x - dx < X \leq x)$$

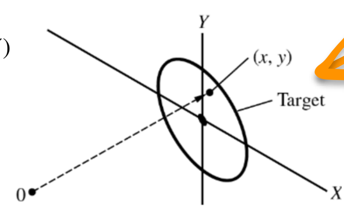
Probability and random variables 7

Random variables and related functions

Joint CDFs and PDFs

- for example a dart thrown at a target – two random variables

$$F_{XY}(x, y) = P(x \leq X, y \leq Y)$$



Target board

- the pdf is defined as $$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2}\int_{x_1}^{x_2} f_{XY}(x, y)dxdy$$

$$F_{XY}(\infty, \infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)dxdy = 1$$

- letting $x_1 = x - dx$, $x_2 = x$, $y_1 = y - dy$ and $y_2 = y$ we obtain

$$f_{XY}(x, y)dxdy = P(x - dx < X \leq x, y - dy < Y \leq y)$$

Probability and random variables 8

Random variables and related functions
Joint CDFs and PDFs

- the cdf for X irrespective of the value of Y is simply

$$F_X(x) = P(X \le x, -\infty < Y \le \infty)$$
$$= F_{XY}(x, \infty)$$

- the cdf for Y alone is $F_Y(y) = F_{XY}(\infty, y)$

- $F_X(x)$ and $F_Y(y)$ are referred to as marginal cdf's and may be expressed as

$$F_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{XY}(x', y')dx'dy' \qquad F_Y(y) = \int_{-\infty}^{y}\int_{-\infty}^{\infty} f_{XY}(x', y')dx'dy'$$

- since $f_X(x) = \dfrac{dF_X(x)}{dx}$ and $f_Y(y) = \dfrac{dF_Y(y)}{dy}$ from

  we obtain $f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y')dy'$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x', y)dx'$

  i.e. the marginal pdf's are obtained by integrating out the undesired variables

Random variables and related functions
Joint CDFs and PDFs

- two random variables are statistically independent if the values each takes on do not influence the values of the other

- for independent random variables, it must be true for any x and y that

$$P(x \le X, y \le Y) = P(x \le X)P(y \le Y)$$

factorize

  or in terms of cdf's $F_{XY}(x, y) = F_X(x)F_Y(y)$

- if we differentiate first w.r.t. x and then y we obtain $f_{XY}(x, y) = f_X(x)f_Y(y)$

- if two random variables are not independent

$$f_{XY}(x, y) = f_X(x)f_{Y|X}(y \mid x)$$
$$= f_Y(y)f_{X|Y}(x \mid y)$$

not factorize

  intuitively $f_{X|Y}(x \mid y)dx = P[x - dx < X \le x \text{ given } Y = y]$

- for independent variables $f_{X|Y}(x \mid y) = f_X(x)$

$$f_{Y|X}(y \mid x) = f_Y(y)$$

Random variables and related functions
Joint CDFs and PDFs

Example

Two random variables X and Y have the joint pdf

$$f_{XY}(x,y) = \begin{cases} Ae^{-(2x+y)}, & x, y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

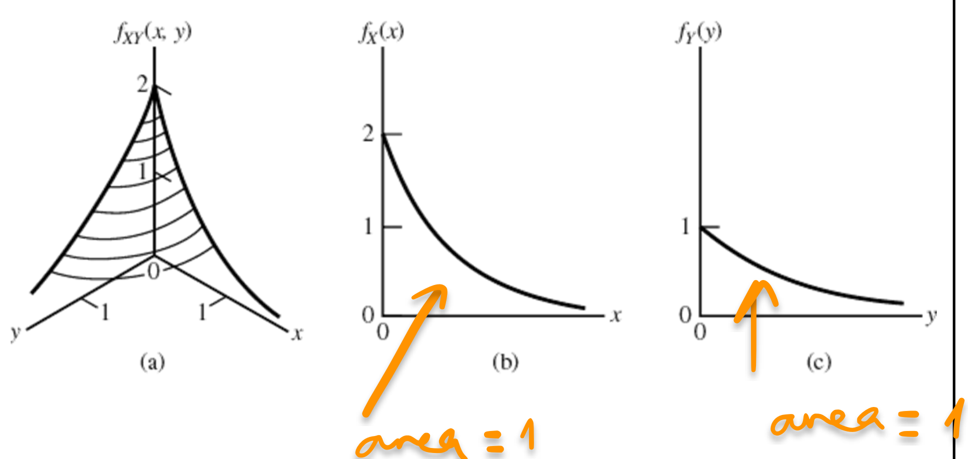
where A is a constant. Determine A and find the two marginal pdf's.

You should see that the rv's are statistically independent.

You should also be able to determine the two cdf's and prove again that the two rv's are independent.

Probability and random variables 11

Random variables and related functions
Joint CDFs and PDFs



Probability and random variables 12

Random variables and related functions

Joint CDFs and PDFs

Example

To illustrate the processes of normalisation of joint pdf's, finding marginal from joint pdf's and checking for statistical independence of the corresponding random variables, we consider the joint pdf

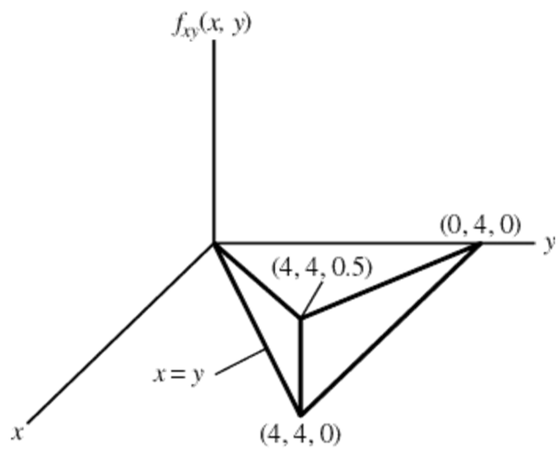
$$f_{XY}(x, y) = \begin{cases} \beta xy, & 0 \le x \le y, 0 \le y \le 4 \\ 0, & \text{otherwise} \end{cases}$$

Determine β and find the two marginal pdf's.

☆ just write down the integrals – pay attention to the limits! They may not be what you first think they are.

Random variables and related functions

Joint CDFs and PDFs



Random variables and related functions

Transformations of random variables

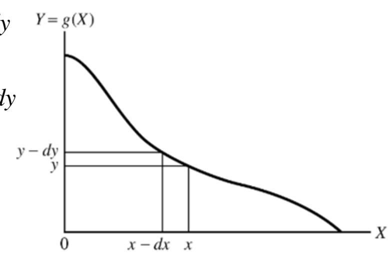
- for situations where we know the pdf of a random variable X and desire the pdf of a second random variable Y defined as a function of X, i.e. $Y = \mathrm{g}(X)$
- consider initially monotonic functions
- the probability that X lies in the range $(x - dx, x)$ is the same as the probability that Y lies in the range $(y - dy, y)$ where $y = g(x)$, thus

$$f_X(x)dx = f_y(y)dy$$

if $g(X)$ is monotonically increasing and

$$f_X(x)dx = -f_y(y)dy$$

if $g(Y)$ is monotonically decreasing



- for both cases $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|_{x=g^{-1}(y)}$

gradient

inverse function

Random variables and related functions

Transformations of random variables

Examples

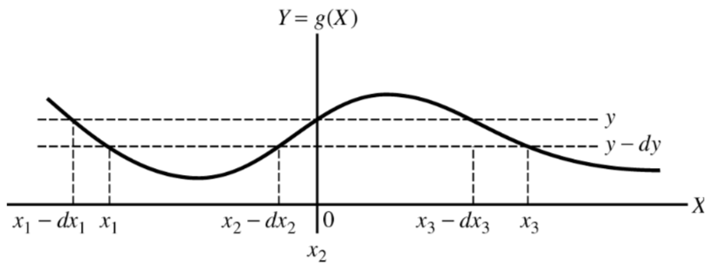
Derive the pdf of the random variable Y defined by

$$Y = -\left(\frac{1}{\pi}\right)\Theta + 1$$

where the random variable Θ has a pdf given by

$$f_\Theta(\theta) = \begin{cases} 1/(2\pi), & 0 \le \theta \le 2\pi \\ 0, & \text{otherwise} \end{cases}$$

Random variables and related functions
Transformations of random variables



- for the case where $g(x)$ is nonmonotonic as shown the infinitesimal interval $(y - dy, y)$ corresponds to three infinitesimal intervals on the x-axis: $(x_1 - dx_1, x_1)$, $(x_2 - dx_2, x_2)$ and $(x_3 - dx_3, x_3)$
- the probability that X lies in any one of these intervals is equal to the probability that Y lies in the interval $(y - dy, y)$

Random variables and related functions
Transformations of random variables

- this can be generalised to the case of N disjoint intervals where it follows that

$$P(y - dy < Y \le y) = \sum_{i=1}^{N} P(x_i - dx_i < X_i \le x_i)$$

where we have generalised to N intervals on the X axis corresponding to the interval $(y - dy, y)$ on the Y axis

- since $P(y - dy < Y \le y) = f_Y(y)dy$ and $P(x_i - dx_i < X_i \le x_i) = f_X(x_i)dx_i$

$$f_Y(y) = \sum_{i=1}^{N} f_X(x_i)\left|\frac{dx_i}{dy}\right|_{x_i = g_i^{-1}(y)}$$

$g_i^{-1}(y)$ is the i^{th} solution to $g(x)=y$

- the absolute value signs insure that probabilities are positive

Random variables and related functions
Transformations of random variables

Example

Consider the transformation $y = x^2$. If $f_X(x) = 0.5\exp(-|x|)$, find $f_Y(y)$.

Random variables and related functions
Transformations of random variables

- suppose two new random variables are defined in terms of two old random variables X and Y by the relations

$$U = g_1(X,Y) \quad \text{and} \quad V = g_2(X,Y)$$

- the new pdf $f_{UV}(u,v)$ is obtained from the old pdf $f_{XY}(x,y)$ by writing

$$P(u-du < U \le u, v-dv < V \le v) = P(x-dx < X \le x, y-dy < Y \le y)$$

or

$$f_{UV}(u,v)dA_{UV} = f_{XY}(x,y)dA_{XY}$$

where dA_{UV} is the infinitesimal area in the uv plane corresponding to the infinitesimal area dA_{XY} in the xy plane through the transformation

- the ratio of elementary area dA_{XY} to dA_{UV} is given by the Jacobian $\dfrac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$

so that $\quad f_{UV}(u,v) = f_{XY}(x,y)\left|\dfrac{\partial(x,y)}{\partial(u,v)}\right|_{\substack{x=g_1^{-1}(u,v)\\ y=g_2^{-1}(u,v)}}$

where the inverse functions $g_1^{-1}(u,v)$ and $g_2^{-1}(u,v)$ exist if we assume a one-to-one transformation

not in Exam

Example

Consider the dart-throwing game discussed in connection with joint cdf’s and and pdf’s. We assume that the joint pdf in terms of rectangular coordinates for the impact point is:

$$f_{XY}(x, y) = \frac{\exp[-(x^2 + y^2)/2\sigma^2]}{2\pi\sigma^2}$$

where σ^2 is a constant. This is a special case of the joint Gaussian pdf. Instead of rectangular coordinates, we wish to use polar coordinates R and Θ, defined by:

$$R = \sqrt{X^2 + Y^2} \qquad \Theta = \tan^{-1}\left(\frac{Y}{X}\right)$$

Determine the pdf of R and Θ.

Statistical averages

- for a discrete random variable

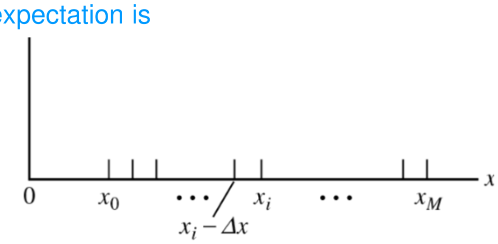
$$\overline{X} = E[X] = \sum_{j=1}^{M} x_j P_j$$

- for a continuous random variable we break up the range of values that X takes into a large number of small sub intervals with length Δx

$$P(x_i - \Delta x < X \le x_i) \cong f_X(x)\Delta x, \quad i = 1,2,\ldots,M$$

  - X is approximated by a discrete random variable that takes on values $x_0, x_1, \ldots x_M$ with probabilities $f_X(x_0), f_X(x_1), \ldots, f_X(x_M)$
  - as Δx approaches dx the expectation is

$$E[X] \cong \sum_{i=0}^{M} x_i f_X(x_i)\Delta x = \int_{-\infty}^{\infty} x\, f_X(x)dx$$



Statistical averages

Average of a function of a random variable

- for a function $y = g(x)$

$$E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)dy$$

where $f_Y(y)$ is the pdf of Y which can be found from $f_X(x)$ from the transformation of a random variable

- sometimes it is more convenient to find the expectation of $g(X)$

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

- the next two examples illustrate this

Statistical averages

Average of a function of a random variable

Example

Suppose the random variable Θ has the pdf

$$f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi}, & |\theta| < \pi \\ 0, & \text{otherwise} \end{cases}$$

Determine $E[\Theta^n]$, referred to as the n^{th} moment of Θ.

The first moment or mean of Θ, $E[\Theta]$, is a measure of the location of $f_\Theta(\theta)$ (i.e. the “centre of mass”). Since $f_\Theta(\theta)$ is symmetrically located about $\theta = 0$, it is not surprising that $E[\Theta] = 0$.

Statistical averages

Average of a function of a random variable

Example

Consider a random variable X that is defined in terms of the uniform random variable Θ considered in the last example by

$$X = \cos\Theta$$

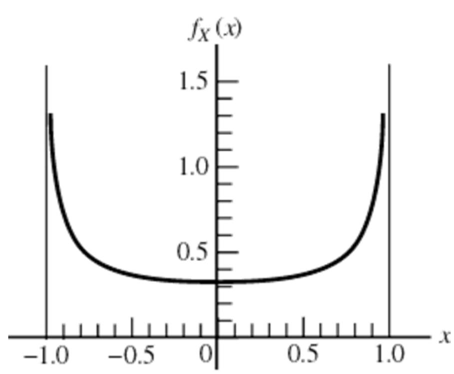
Determine the density function of X, $f_X(x)$ and the first and second moments.

You will need to use: $\dfrac{d}{dx}\cos^{-1} x = \dfrac{1}{\sqrt{1-x^2}}$

and: $\displaystyle\int_{-1}^{1} \frac{x^2}{\pi\sqrt{1-x^2}}dx = \frac{1}{2}$ … unless … ?

Statistical averages

Average of a function of a random variable



Statistical averages

Average of a function of multiple random variables

marginalisation

- if $f_{XY}(x,y)$ is the joint pdf of X and Y the expectation of $g(X,Y)$ is

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{XY}(x,y)dxdy$$

  and the generalisation to more than two random variables is obvious

- note that if $g(X,Y)$ is replaced by a function of X alone, say $h(X)$, we obtain

$$E[h(X)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)\, f_{XY}(x,y)dxdy$$
$$= \int_{-\infty}^{\infty} h(x)\, f_X(x)dx$$

another marginal

- the concept of conditional expectation may be easier, e.g. for a function $g(X,Y)$ of two random variables X and Y with the joint pdf $f_{XY}(x,y)$

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, f_{XY}(x,y)dxdy$$
$$= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y) f_{X|Y}(x\,|\,y)dx\right] f_Y(y)dy$$
$$= E\{E[g(X,Y)\,|\,Y]\}$$

marginal pdf

Example

Consider the joint pdf

$$f_{XY}(x,y) = \begin{cases} Ae^{-(2x+y)}, & x,y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Noting that X and Y are statistically independent, determine the expectation of $g(X, Y) = X\,Y$.

☆ it's the same example as earlier!

Statistical averages

Average of a function of multiple random variables

Example

As a specific example of conditional expectation, consider the firing of projectiles at a target. Projectiles are fired until the target is hit for the first time, after which firing ceases. Assume that the probability of a projectile's hitting the target is p and that the firings are independent of one another. Find the average number of projectiles fired at the target.

Statistical averages

Variance of a random variable

- the variance, $\mathrm{var}\{X\}$ or σ_X^2 is given by

$$\sigma_X^2 = E\{[X - E(X)]^2\}$$

- the standard deviation σ_X measures the concentration of the pdf of X, or $f_X(x)$, about the mean

- a useful relation for obtaining σ_X^2 is

$$\sigma_X^2 = E[X^2] - E^2[X]$$

which is the second moment minus the mean squared

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x)dx = \int_{-\infty}^{\infty} (x^2 - 2xm_X + m_X^2) f_X(x)dx$$
$$= E[X^2] - 2m_X^2 + m_X^2 = E[X^2] - E^2[X]$$

which follows since $\int_{-\infty}^{\infty} x\, f_X(x)dx = m_X$

Statistical averages
Variance of a random variable

Example

Determine the variance of the uniform pdf

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

Statistical averages
Average of a linear combination of N random variables

- the expected value of a linear combination of random variables is the same as the linear combination of their respective means

linear combination

$$E\left[\sum_{i=1}^{N} a_i X_i\right] = \sum_{i=1}^{N} a_i E[X_i]$$

where $X_1, X_2, \ldots, X_N$ are random variables and $a_1, a_2, \ldots, a_N$ are arbitrary constants

- demonstrating for the case where $N = 2$ and $f_{X1X2}(x_1,x_2)$, the joint pdf of X_1 and X_2

$$\begin{aligned} E[a_1 X_1 + a_2 X_2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (a_1 x_1 + a_2 x_2) f_{X_1 X_2}(x_1, x_2)\, dx_1 dx_2 \\ &= a_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f_{X_1 X_2}(x_1, x_2)\, dx_1 dx_2 \\ &\quad + a_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f_{X_1 X_2}(x_1, x_2)\, dx_1 dx_2 \end{aligned}$$

Statistical averages

Average of a linear combination of N random variables

- considering the first double integral we find that

another marginal

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f_{X_1X_2}(x_1,x_2)dx_1dx_2 = \int_{-\infty}^{\infty} x_1\left\{\int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)dx_2\right\}dx_1$$

$$= \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)dx_1 = E[X_1]$$

- similarly it can be shown that the second double integral equals $E[X_2]$
- the result holds for any N
- does the result hold for dependent and independent variables?

Statistical averages

Variance of a linear combination of independent random variables

- if $X_1, X_2, \ldots, X_N$ are statistically independent random variables then

$$\mathrm{var}\left\{\sum_{i=1}^{N} a_i X_i\right\} = \sum_{i=1}^{N} a_i^2 \,\mathrm{var}\{X_i\}$$

where $a_1, a_2, \ldots, a_N$ are arbitrary constants and $\mathrm{var}\{X_i\} = E[(X_i - \overline{X_i})^2]$

- again demonstrating the case for $N = 2$
  - let $Z = a_1X_1+a_2X_2$ and $f_{Xi}(x_i)$ be the marginal pdf of X_i
  - the joint pdf of X_1 and X_2 is $f_{X1}(x_1)\,fX_2(x_2)$ due to independence
  - in addition $\overline{Z} = a_1\overline{X_1} + a_2\overline{X_2}$ and $\mathrm{var}\{Z\} = E[(Z-\overline{Z})^2]$
  - since $Z=a_1X_1+a_2X_2$ we may write $\mathrm{var}\{Z\}$ as

$$\mathrm{var}\{Z\} = E\left\{[(a_1X_1 + a_2X_2) - (a_1\overline{X_1} + a_2\overline{X_2})]^2\right\}$$

$$= E\left\{[a_1(X_1 - \overline{X_1}) + a_2(X_2 - \overline{X_2})]^2\right\}$$

$$= a_1^2E[(X_1-\overline{X_1})^2] + 2a_1a_2E[(X_1-\overline{X_1})(X_2-\overline{X_2})] + a_2^2E[(X_2-\overline{X_2})^2]$$

Statistical averages

Variance of a linear combination of independent random variables

$$\begin{aligned}\text{var}\{Z\} &= E\left\{[(a_1X_1 + a_2X_2) - (a_1\overline{X_1} + a_2\overline{X_2})]^2\right\} \\ &= E\left\{[a_1(X_1 - \overline{X_1}) + a_2(X_2 - \overline{X_2})]^2\right\} \\ &= a_1^2 E[(X_1 - \overline{X_1})^2] + 2a_1a_2E[(X_1 - \overline{X_1})(X_2 - \overline{X_2})] + a_2^2 E[(X_2 - \overline{X_2})^2]\end{aligned}$$

- the first and last terms are $\text{var}\{X_1\}$ and $\text{var}\{X_2\}$
- the middle term is zero since

$$\begin{aligned}E[(X_1 - \overline{X_1})(X_2 - \overline{X_2})] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(X_1 - \overline{X_1})(X_2 - \overline{X_2}) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1 dx_2 \\ &= \int_{-\infty}^{\infty}(X_1 - \overline{X_1}) f_{X_1}(x_1)\, dx_1 \int_{-\infty}^{\infty}(X_2 - \overline{X_2}) f_{X_2}(x_2)\, dx_2 \\ &= (\overline{X_1} - \overline{X_1})(\overline{X_2} - \overline{X_2}) = 0\end{aligned}$$

Probability and random variables 35

Statistical averages

The characteristic function

- letting $g(X) = \exp(jvX)$ we obtain the characteristic function of X

$$M_X(jv) = E[e^{jvX}] = \int_{-\infty}^{\infty} f_X(x) e^{jvx} dx$$

- with jv in the exponent replaced by $-j\omega$, $M_X(jv)$ would be the Fourier transform of $f_X(x)$
- $f_X(x)$ is obtained from $M_X(jv)$ according to the inverse transform

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(jv) e^{-jvx} dv$$

which is useful when the pdf of a random variable is sought, but the characteristic function is more easily obtained

Probability and random variables 36

Statistical averages
The characteristic function

- the characteristic function may be used to obtain the moments of a random variable
- if we differentiate $M_X(jv)$ w.r.t. v

$$\frac{\partial M_X(jv)}{\partial v} = j\int_{-\infty}^{\infty} x f_X(x) e^{jvx} dx$$

- setting $v = 0$ and dividing by j, we obtain

$$E[X] = (-j)\frac{\partial M_X(jv)}{\partial v}\bigg|_{v=0}$$

and by repeated differentiation

$$E[X^n] = (-j)^n \frac{\partial^n M_X(jv)}{\partial v^n}\bigg|_{v=0}$$

Statistical averages
The characteristic function

Example

Use a table of Fourier transforms to obtain the characteristic function of the one-sided exponential pdf

$$f_X(x) = \exp(-x)u(x)$$

and determine an expression for its n^{th} moment.

$$\exp(-at)u(t) \leftrightarrow \frac{1}{a + j2\pi f}$$

Statistical averages

The PDF of the sum of two independent random variables

- we can use the characteristic function to determine the pdf of a sum of two independent random variables X & Y, i.e. $Z = X + Y$

$$M_Z(jv) = E[e^{jvZ}] = E[e^{jv(X+Y)}]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{jv(x+y)} f_X(x) f_Y(y) dxdy$$

since the joint pdf of X and Y is $f_X(x)f_Y(y)$ due to independence

- we can write this expression as the product of two integrals since $\exp(jv[x+y]) = \exp(jvx)\exp(jvy)$

$$M_Z(jv) = \int_{-\infty}^{\infty} f_X(x)e^{jvx}dx \int_{-\infty}^{\infty} f_Y(y)e^{jvy}dy$$
$$= E[e^{jvX}]E[e^{jvY}]$$

Probability and random variables 39

Statistical averages

The PDF of the sum of two independent random variables

- from the definition of the characteristic function we see that

$$M_Z(jv) = M_X(jv)M_Y(jv)$$

- remembering the similarity to the Fourier transform and that a product in the frequency domain corresponds to convolution in the time domain

$$f_Z(z) = f_X(x) * f_Y(y) = \int_{-\infty}^{\infty} f_X(z-u) f_Y(u) du$$

Probability and random variables 40

Example

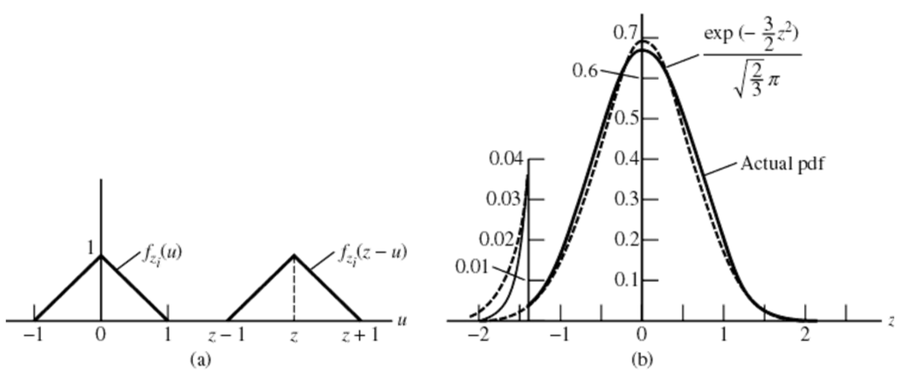
First sketch and then determine the pdf of Z, the sum of four identically distributed, independent random variables,

$$Z = X_1 + X_2 + X_3 + X_4$$

where the pdf of each X_i is given by

$$f_{X_i}(x_i) = \prod(x_i) = \begin{cases} 1, & |x_i| \le \frac{1}{2} \\ 0, & \text{otherwise}, i = 1,2,3,4 \end{cases}$$

and where $\prod(x_i)$ is the unit rectangular pulse function.



Statistical averages
Covariance and correlation coefficients

- the covariance of two random variables X and Y is defined by

$$\mu_{XY} = E[(X-\bar{X})(Y-\bar{Y})] = E[XY] - E[X]E[Y]$$

- the correlation coefficient is defined by

$$\text{rho} \longrightarrow \rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y} \Rightarrow [-1, 1]$$

0 uncorrelated

- thus we have the relationship

$$E[XY] = \sigma_X \sigma_Y \rho_{XY} + E[X]E[Y]$$

- both ρ_{XY} and μ_{XY} are measures of the interdependence of X and Y

- the normalisation of the correlation coefficient is such that $-1 \le \rho_{XY} \le 1$

Probability and random variables 43

Statistical averages
Covariance and correlation coefficients

- if X and Y are independent, their joint pdf $f_{XY}(x,y)$ is the product of the two respective marginal pdfs, that is $f_{XY}(x,y) = f_X(x)f_Y(y)$, thus

$$\begin{aligned}
\mu_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{X})(y-\bar{Y})f_X(x)f_Y(y)\,dxdy \\
&= \int_{-\infty}^{\infty}(x-\bar{X})f_X(x)dx\int_{-\infty}^{\infty}(y-\bar{Y})f_Y(y)dy \\
&= (\bar{X}-\bar{X})(\bar{Y}-\bar{Y}) = 0
\end{aligned}$$

- considering the case where $X = \pm\alpha Y$, where α is a positive constant

$$\begin{aligned}
\mu_{XY} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\pm\alpha y \mp \alpha\bar{Y})(y-\bar{Y})f_{XY}(x,y)\,dxdy \\
&= \pm\alpha\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\bar{Y})^2 f_{XY}(x,y)\,dxdy \\
&= \pm\alpha\sigma_Y^2
\end{aligned}$$

- we can write the variance of X as $\sigma_X^2 = \alpha^2\sigma_Y^2$, thus

$$\rho_{XY} = +1, \text{ for } X = +\alpha Y \text{ and } \rho_{XY} = -1, \text{ for } X = -\alpha Y, \ \alpha > 0$$

Probability and random variables 44

Some useful PDFs

- binomial
- Poisson
- geometric
- Gaussian
  - the central limit theorem
  - the Gaussian Q-function
- Rayleigh
- exponential

Binomial distribution

- consider a chance experiment where there are two mutually exclusive, exhaustive outcomes A and $\bar{A}$ with probabilities $p(A) = p$ and $p(\bar{A}) = q = 1 - p$

- K is the number of times event A occurs in N trials and we seek the probability that exactly $k \leq n$ occurrences of the event A occur in n repetitions of the experiment

- e.g. consider a coin tossing experiment where we wish to obtain the probability of k heads in n tosses of the coin
  - $p(\text{head}) = p$ and $p(\text{tail}) = q = 1 - p$

- one possible sequence of k heads in n tosses is

$$\underbrace{HH\ldots H}_{k \text{ heads}}\underbrace{TT\ldots T}_{n-k \text{ tails}}$$

Some useful PDFs
Binomial distribution

- since the trials are independent, the probability of this sequence is

$$\underbrace{p \cdot p \cdot p \cdot \ldots p}_{k \text{ factors}} \cdot \underbrace{q \cdot q \cdot q \cdot \ldots q}_{n-k \text{ factors}} = p^n q^{n-k}$$

- but this is only one sequence which yields this number of heads and tails – in total there are

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

sequences with the same number of heads and tails and probability

- the binomial coefficient comes from considering the number of different ways k **identifiable** heads can be arranged in n slots
  - the first head can fall in any of the n slots, the second in any of $n-1$ slots,…

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Some useful PDFs
Binomial distribution

- but we aren't concerned about which head occupies which slot and there are $k!$ arrangements
  - the total number of arrangements is

$$\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

- since the occurrence of any of these possible arrangements precludes the occurrence of any other, and since each occurs with probability $p^k q^{n-k}$, the probability of *exactly* k heads in n trials in **any** order is

$$P(K=k) = P_n(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0,1,\ldots,n$$

- this is the binomial probability distribution

Some useful PDFs
Binomial distribution

- the mean of a binomial distributed random variable, K, is given by

$$E[K]=\sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

which can be started at $k = 1$ since the first term is zero

$$E[K]=\sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k q^{n-k}$$

- letting $m = k - 1$

$$E[K]=\sum_{m=0}^{n-1} \frac{n!}{m!(n-m-1)!}p^{m+1} q^{n-m-1}$$

$$=np\sum_{m=0}^{n-1} \frac{(n-1)!}{m!(n-m-1)!}p^{m} q^{n-m-1}$$

Probability and random variables 49

Some useful PDFs
Binomial distribution

- letting $\ell = n - 1$ and recalling that, by the binomial theorem

$$(x+y)^{\ell}=\sum_{m=0}^{\ell}\binom{\ell}{m}x^m y^{\ell-m}$$

- we obtain

$$\overline{K}=E[K]=np(p+q)^{\ell}=np$$

since $p + q = 1$

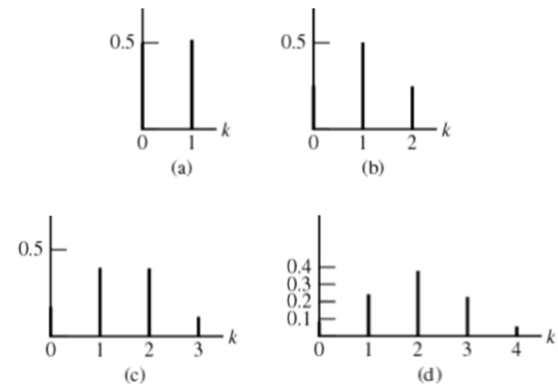
- thus in a long sequence of n tosses of a coin we would expect about $np = ½n$ heads

- a similar series of manipulations shows that $E[K^2] = np(np + q)$
  - the variance of a binomial distributed random variable is thus

$$\sigma_K^2=E[K^2]-E^2[K]=npq=\overline{K}(1-p)$$

Probability and random variables 50

Some useful PDFs
Binomial distribution



The binomial distribution

(a) $n = 1$, $p = 0.5$. (b) $n = 2$, $p = 0.5$. (c) $n = 3$, $p = 0.5$. (d) $n = 4$, $p = 0.5$.

Some useful PDFs
Binomial distribution

Example

Assuming single births, determine the probability of having 1, 2, 3 and 4 girls in a four-child family when the probability of a female is 0.5.

Some useful PDFs

Poisson distribution

- consider a chance experiment in which an event whose probability of occurrence in a very small time interval ΔT is $P = \alpha\,\Delta T$, where α is a constant of proportionality
- if successive occurrences are statistically independent, then the probability of k events in time T is
  $$P_T(k) = \frac{(\alpha T)^k}{k!} e^{-\alpha T}, \quad k \text{ integer}$$
  - this is the Poisson distribution
- the Poisson distribution can be used to approximate the binomial distribution when the number of trials n is large, the probability of each event p is small, and the product $np \approx npq$
  $$P_n(k) \cong \frac{(\overline{K})^k}{k!} e^{-\overline{K}}, \quad k \text{ integer}$$
- where $\overline{K} = E[K] = np$ and $\sigma_k^2 = E[K]q = npq \approx E[K]$ for $q = 1 - p \approx 1$

Some useful PDFs

Poisson distribution

Example

The probability of error on a single transmission in a digital communication system is $P_E = 10^{-4}$. Using the Poisson approximation to the binomial distribution determine the probability of more than three errors in 1000 transmissions?

Some useful PDFs

Geometric distribution

- examples:
  - the probability of the first head in a series of coin tossings, or
  - the first error in a long string of digital signal transmissions

  occurring on the k^{th} trial

- these can be described by the geometric distribution

$$P(k) = pq^{k-1}, \quad 1 \le k < \infty$$

where p is the probability of the event of interest and q is the probability of it not occurring

Some useful PDFs

Geometric distribution

Example

What is the probability of the first error occurring at the 1000$^{\text{th}}$ transmission in a digital data transmission system where the probability of error is $p = 10^{-6}$?

Some useful PDFs
Central limit theorem

- $X_1, X_2, \ldots$ are independent, identically distributed random variables, each with finite mean m and finite variance σ^2

- Z_n is a sequence of unit-variance, zero-mean random variables defined as

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}}$$

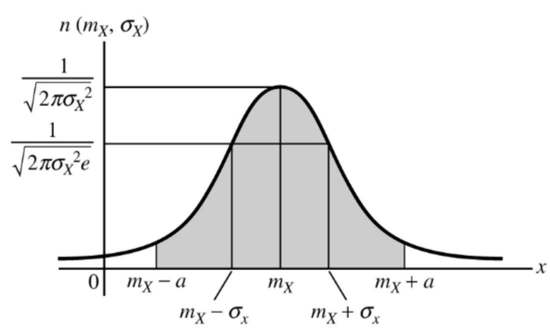
then

$$\lim_{n\to\infty} P(Z_n \le z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\, dt$$

i.e. the cdf of the **normalised** sum approaches a Gaussian cdf, no matter what the distribution of the component random variables

Some useful PDFs
Gaussian distribution

- you will undoubtedly make much use of the Gaussian distribution
  - assumption of Gaussian statistics for random phenomena can make intractable problem tractable
  - according to the central limit theorem many naturally occurring random quantities such as noise or measurement errors are Gaussian distributed



Some useful PDFs
Gaussian distribution

- the generalisation of the joint Gaussian pdf is

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left\{\frac{\left(\frac{x-m_x}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-m_x}{\sigma_X}\right)\left(\frac{y-m_y}{\sigma_Y}\right) + \left(\frac{y-m_y}{\sigma_Y}\right)^2}{2(1-\rho^2)}\right\}$$

where we can show that

$$m_X = E[X] \quad \text{and} \quad m_Y = E[Y]$$

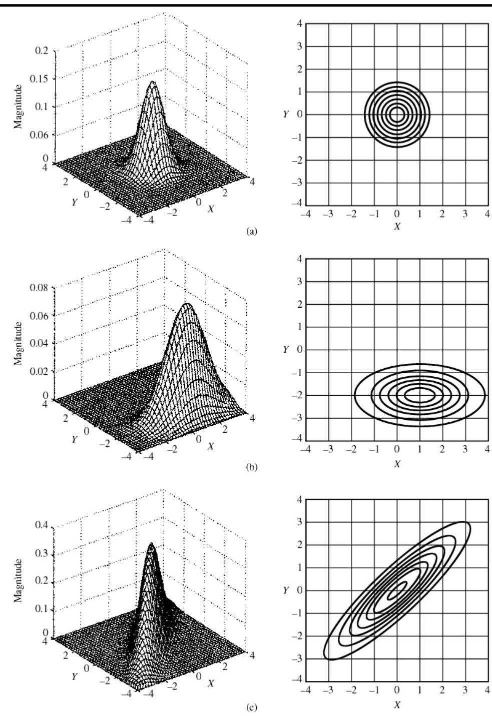
$$\sigma_X^2 = \mathrm{var}\{X\}$$

$$\sigma_Y^2 = \mathrm{var}\{Y\}$$

and

$$\rho = \frac{E[(X-m_X)(Y-m_Y)]}{\sigma_X\sigma_Y}$$

$m_X = 0, m_Y = 0$
$\sigma_X^2 = 1, \sigma_Y^2 = 1$
$\rho = 0$



$m_X = 1, m_Y = -2$
$\sigma_X^2 = 2, \sigma_Y^2 = 1$
$\rho = 0$

$m_X = 0, m_Y = 0$
$\sigma_X^2 = 1, \sigma_Y^2 = 1$
$\rho = 0.9$

Some useful PDFs
Gaussian distribution

- the marginal pdf for X (or Y) can be obtained by integrating over y (or x) – for X it is given by

$$n(m_X, \sigma_X) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-(x-m_X)^2/(2\sigma_X^2)\right]$$

where $n(m_X, \sigma_X)$ is the notation for a Gaussian random variable with mean m_X and variance σ_X

- for $\rho = 0$, i.e. X & Y uncorrelated, the general joint Gaussian pdf becomes

$$f_{XY}(x,y) = \frac{\exp(x^2/(2\sigma_X^2))}{\sqrt{2\pi\sigma_X^2}} \frac{\exp(y^2/(2\sigma_Y^2))}{\sqrt{2\pi\sigma_Y^2}} = f_X(x) f_Y(y)$$

thus uncorrelated Gaussian random variables are also statistically independent – this does not hold for all pdfs

Probability and random variables 61

Some useful PDFs
Gaussian distribution

- for $Z = X_1 + X_2$ where X_1 and X_2 are two independent Gaussian random variables with pdf $n(m_i, \sigma_i)$

$$M_{X_i}(jv) = \int_{-\infty}^{\infty} (2\pi\sigma_i^2)^{-1/2} \exp\left[\frac{-(x_i-m_i)^2}{2\sigma_i^2}\right] \exp(jvx_i)\,dx_i$$

$$= \exp\left(jm_i v - \frac{\sigma_i^2 v^2}{2}\right)$$

which is obtained from a table of Fourier transforms or by completing the square and integrating

- thus the characteristic function of Z is

$$M_Z(jv) = M_{X_1}(jv) M_{X_2}(jv) = \exp\left[j(m_1+m_2)v - \frac{(\sigma_1^2 - \sigma_2^2)v^2}{2}\right]$$

which is the characteristic function of a Gaussian random variable of mean $m_1 + m_2$ and variance $\sigma_1^2 + \sigma_2^2$

Probability and random variables 62

Some useful PDFs
Gaussian *Q*-function

- the symmetry of the Gaussian distribution ensures that

$$P(X \le m_X) = P(X \ge m_X) = 0.5$$

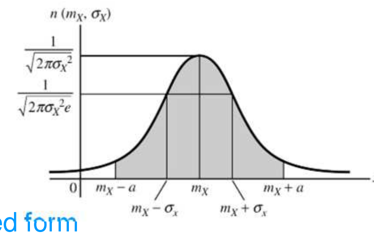
- suppose we wish to find

$$P[m_X - a \le X \le m_X + a] = \int_{m_X - a}^{m_X + a} \frac{\exp\left[-(x-m_X)^2/2\sigma_X^2\right]}{\sqrt{2\pi\sigma_X^2}} dx$$

changing variables give us

$$P[m_X - a \le X \le m_X + a] = \int_{-a/\sigma_X}^{a/\sigma_X} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

$$= 2\int_0^{a/\sigma_X} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$



- the integral cannot be evaluated in closed form

Some useful PDFs
Gaussian *Q*-function

- the Gaussian *Q*-function is defined as

$$Q(u) = \int_u^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$$

- the function has been evaluated numerically and rational and asymptotic approximations are available to evaluate it for moderate and large arguments

- we may write

$$P[m_X - a \le X \le m_X + a] = 2\left[\frac{1}{2} - \int_{a/\sigma_X}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy\right]$$

$$= 1 - 2Q\left(\frac{a}{\sigma_X}\right)$$

and a useful approximation to $Q(u)$ is given by $Q(u) \cong \dfrac{e^{-u^2/2}}{u\sqrt{2\pi}}, \quad u >> 1$

Some useful PDFs
Gaussian Q-function

- there is less than 6 % error for $u \geq 3$
- related integrals are the error function and the complementary error function defined as

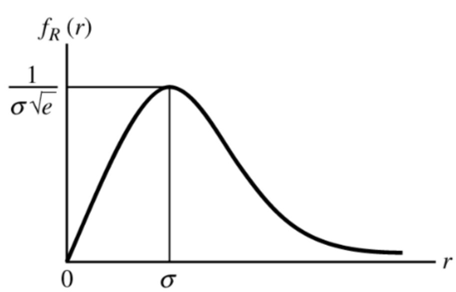
$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-y^2}\, dy$$

$$\operatorname{erfc}(u) = 1 - \operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_u^\infty e^{-y^2}\, dy$$

- useful relationships to the Q-function

$$\operatorname{erf}(u) = 1 - 2Q(\sqrt{2}u) \quad \text{or} \quad Q(\sqrt{2}u) = \frac{1}{2}\operatorname{erfc}(u)$$

Some useful PDFs
Rayleigh distribution



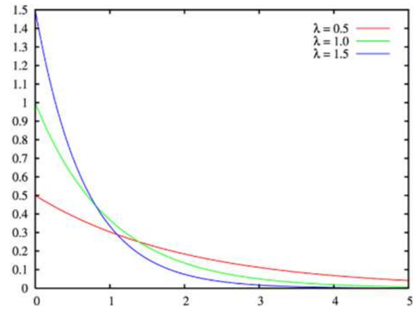
$$f_R(r) = \frac{r}{\sigma^2} \exp\left(\frac{-r^2}{2\sigma^2}\right)$$

- we've already seen the Rayleigh pdf in a previous example
- we will consider another example, more related to communications, in the tutorial

Some useful PDFs

Exponential distribution

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



- we will see an example of the exponential distribution in the tutorial