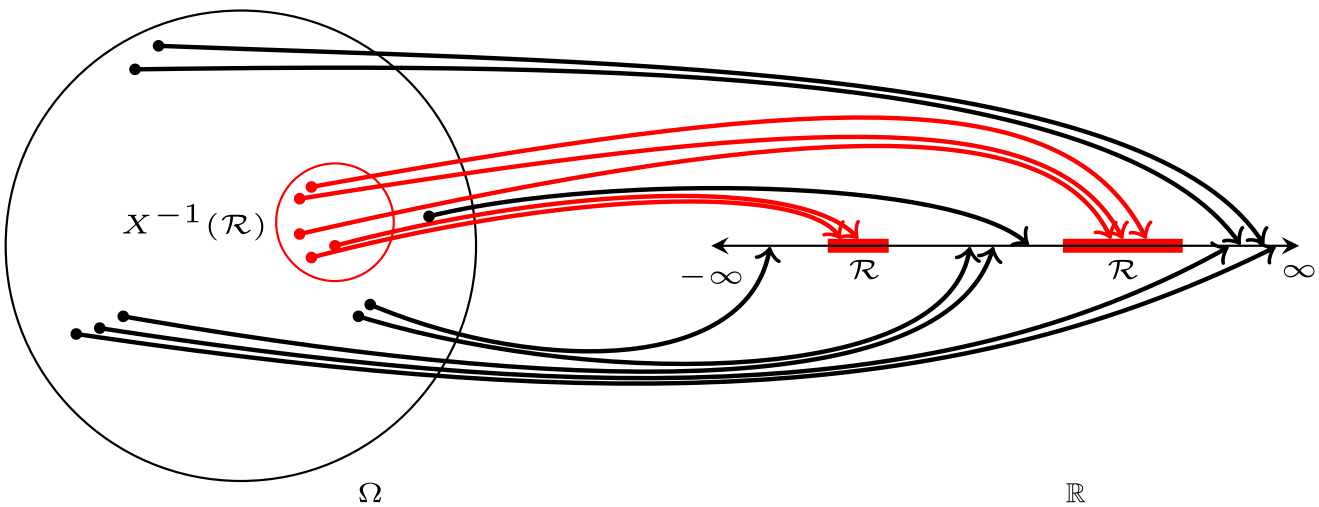
Chapter 2

Random Variables

Consider a probability space $(\Omega, \mathcal{F}, P)$. A *random variable*, X, is a function from the sample space Ω to the real line $\mathbb{R}$ which associates to each outcome ω a real value $X(\omega)$, such that for any subset $\mathcal{R} \subseteq \mathbb{R}$, obtained by taking unions, intersections, and complements of a discrete number (finite or infinite) of open or closed intervals in $\mathbb{R}$, the set $X^{-1}(\mathcal{R}) = \{\omega : X(\omega) \in \mathcal{R}\}$ is an event in $\mathcal{F}$. We use $P(X \in \mathcal{R})$ to denote the probability of this event, i.e.,

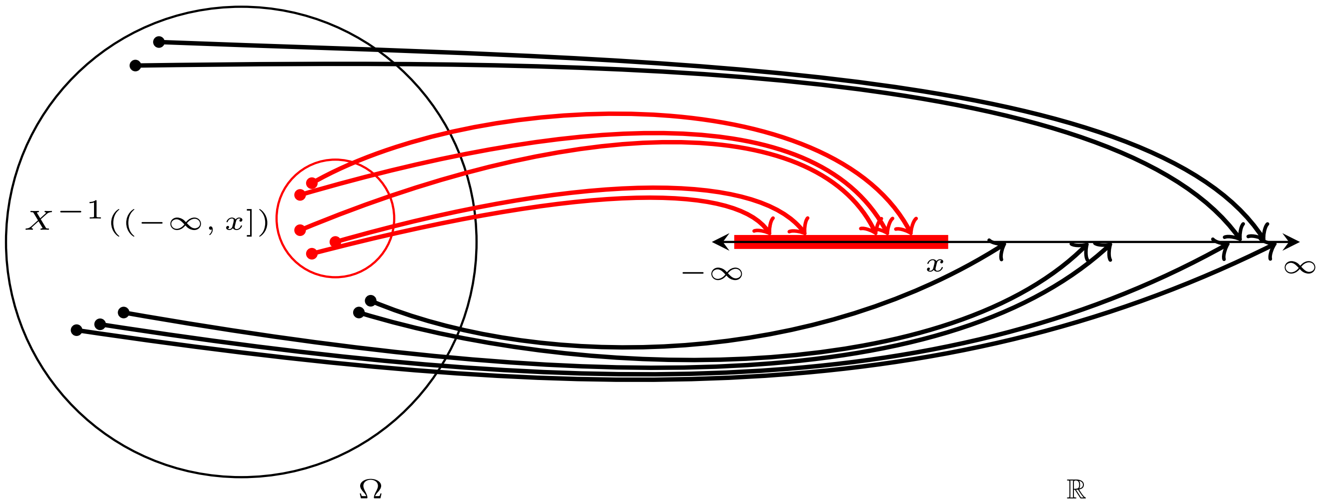
$$P(X \in \mathcal{R}) = P(X^{-1}(\mathcal{R})) = P(\{\omega : X(\omega) \in \mathcal{R}\}).$$



In particular, for any real number x,

$$X^{-1}((-\infty, x]) = \{\omega : X(\omega) \leq x\}$$

is an event with probability $P(X \leq x)$. This probability, denoted by $F_X(x)$, is called the *cumulative distribution function* (CDF) of X.



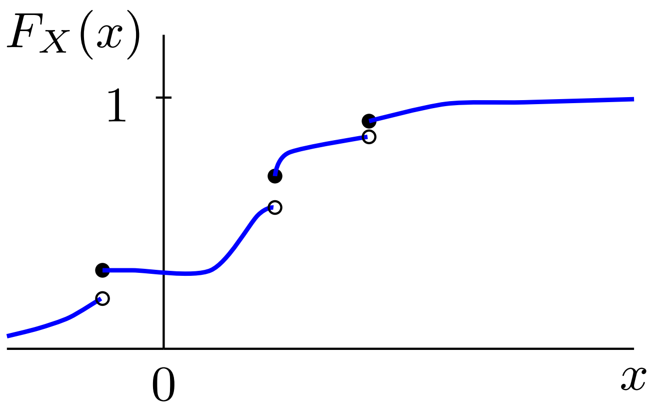
Properties of the Cumulative Distribution Function

i) $F_X(-\infty) = 0$, $F_X(\infty) = 1$.

ii) $F_X(x)$ is nondecreasing, i.e., $F_X(x_1) \leq F_X(x_2)$ for $x_1 \leq x_2$.

iii) $F_X(x)$ is right continuous, i.e., $\lim_{x \searrow x_0} F_X(x) = F_X(x_0)$.

iv) Let $F_X(x_0^-) = \lim_{x \nearrow x_0} F_X(x)$. Then, $P(X = x_0) = F_X(x_0) - F_X(x_0^-)$. Thus, X takes the value x_0 with nonzero probability if and only if $F_X(x)$ is discontinuous at x_0.

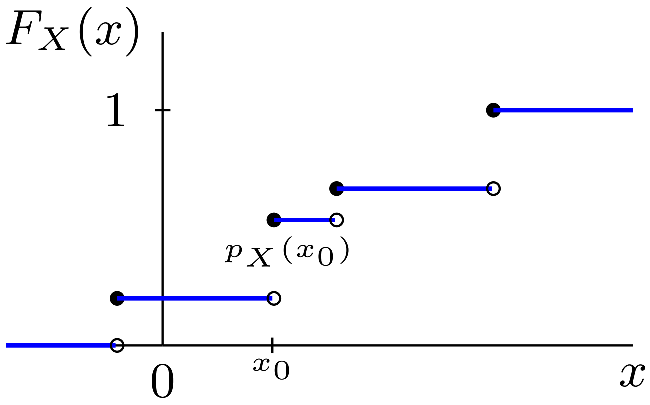


Discrete Random Variables

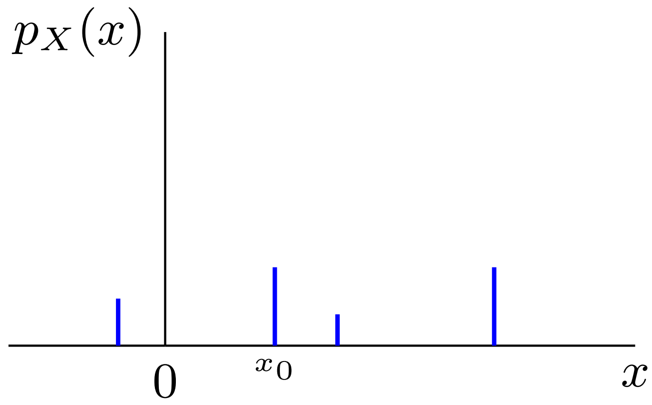
A random variable X is *discrete* if it takes a discrete number of values. The CDF of X is a staircase function with discontinuities at the values assumed by X. The probability that X assumes the value x_0 is

$$p_X(x_0) = P(X = x_0) = P(\{\omega : X(\omega) = x_0\}) = F_X(x_0) - F_X(x_0^-).$$

This function is called the *probability mass function* (PMF) of the random variable X.



CDF



PMF

Continuous Random Variables

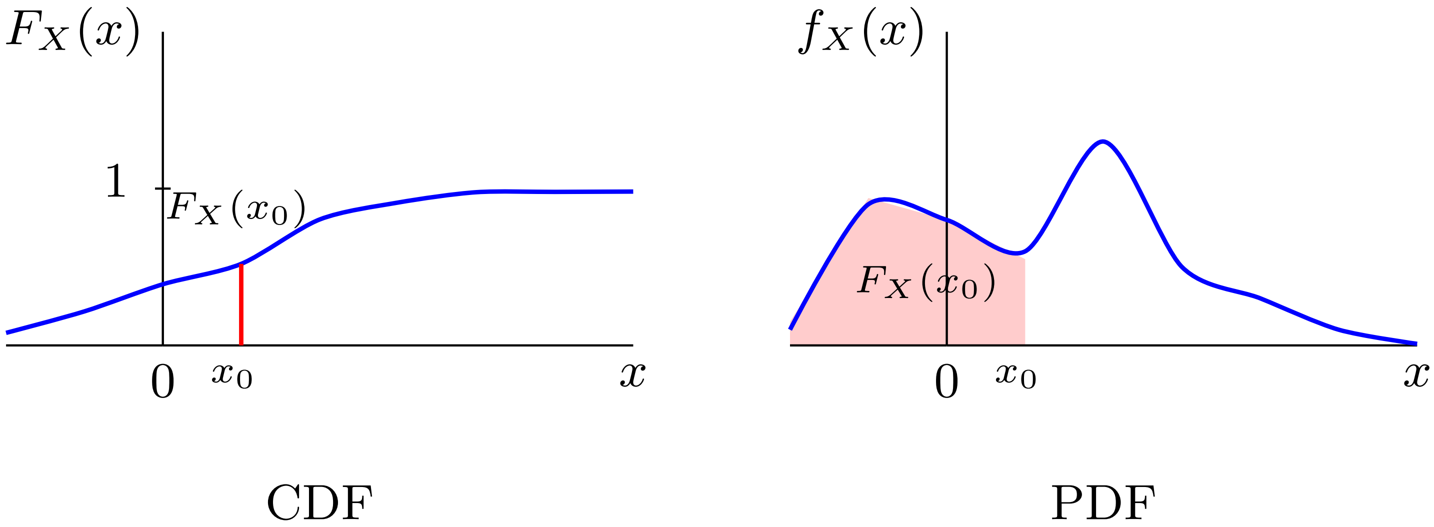
A random variable X is *continuous* if there is a function $f_X(x)$ such that for all $x_0 \in \mathbb{R}$,

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$

The function $f_X(x)$ is called the *probability density function* (PDF) of the random variable X. In particular, a continuous random variable has a continuous cumulative distribution function.

Since $F_X(x)$ is nondecreasing, $f_X(x)$ is nonnegative. Furthermore, since $F_X(\infty) = 1$, we have $\int_{-\infty}^{\infty} f_X(x)dx = 1$.

If $F_X(x)$ is differentiable, then $f_X(x)$ is its derivative.


$F_X(x)$
1
$F_X(x_0)$
0
x_0
x
CDF
$f_X(x)$
$F_X(x_0)$
0
x_0
x
PDF

The *conditional CDF* of a random variable X given an event E, with $P(E) > 0$, is given by

$$F_X(x|E) = P(X \leq x|E) = \frac{P(X \leq x \cap E)}{P(E)}$$

and the *conditional PDF*, $f_X(x|E)$, of X given E is such that

$$F_X(x|E) = \int_{-\infty}^{x} f_X(x|E)dx$$

for all x. Hence, for small h,

$$\begin{aligned} P(x < X \leq x+h|E) &= P(X \leq x+h|E) - P(X \leq x|E) \\ &= F_X(x+h|E) - F_X(x|E) \\ &= \int_{x}^{x+h} f_X(x|E)dx \approx hf_X(x|E). \end{aligned}$$

We conclude that $f_X(x|E)$ is the limit of $P(x < X \leq x+h|E)/h$ as $h \to 0$ provided that the limit exists.

Function of a Random Variable

Let X be a random variable and $Y = g(X)$, for some function $g : \mathbb{R} \to \mathbb{R}$. Then, Y is a random variable.

Let y be a real number and $\mathcal{S}_{\leq y} = \{x : g(x) \leq y\}$. Then,

$$F_Y(y) = P(Y \leq y) = P(X \in \mathcal{S}_{\leq y}).$$

PMF of a Function of a Discrete Random Variable

If X is discrete, then $Y = g(X)$ is discrete.
Let $\mathcal{S}_{=y} = \{x : g(x) = y\}$. We have $p_Y(y) = P(X \in \mathcal{S}_{=y})$.

Example 1 Let X be a discrete random variable assuming the values $-2, -1, 0, 1, 2$ with probabilities $\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}$, respectively. Then, $Y = X^2$ is a discrete random variable which assumes the values $0, 1, 4$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, respectively.

CDF of a Function of a Continuous Random Variable

Example 2 Let $Y = X^2$. Then, for $y < 0$, $\mathcal{S}_{\leq y}$ is empty and $F_Y(y) = 0$. For $y \geq 0$, $\mathcal{S}_{\leq y} = [-\sqrt{y}, \sqrt{y}]$. Hence,

$$
\begin{aligned}
F_Y(y) &= P(X \in [-\sqrt{y}, \sqrt{y}]) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}).
\end{aligned}
$$

In summary,

$$
F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y \geq 0. \end{cases}
$$

PDF of a Function of a Continuous Random Variable

Let X be a continuous random variable and $Y = g(X)$. If $F_Y(y)$ is differentiable for all y, then Y is a continuous random variable and we can find its PDF by differentiating $F_Y(y)$.

Example 3 Let $Y = X^2$, we have shown in Example 2 that

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y \geq 0. \end{cases}$$

Differentiating we get,

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}} \left(f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right) & y \geq 0. \end{cases}$$

Another method is to find the PDF of Y directly from that of X. Let $x_1 = h_1(y), x_2 = h_2(y), \ldots$ be all values of x for which $g(x) = y$. Suppose further that for each such x_i, the derivative $g'(x_i)$ exists and is nonzero. Then,

$$f_Y(y) = \sum_i \left. \frac{f_X(x_i)}{|g'(x_i)|} \right|_{x_i = h_i(y)} .$$

Example 4 In Example 2, with $Y = g(X) = X^2$, let $y > 0$. Then, there are two values $x_1 = \sqrt{y}$ and $x_2 = -\sqrt{y}$ for which $x_i^2 = y$. We have $g'(x_1) = 2x_1 = 2\sqrt{y}$ and $g'(x_2) = 2x_2 = -2\sqrt{y}$. Hence,

$$
\begin{aligned}
f_Y(y) &= \left.\frac{f_X(x_1)}{|g'(x_1)|}\right|_{x_1=\sqrt{y}} + \left.\frac{f_X(x_2)}{|g'(x_2)|}\right|_{x_2=-\sqrt{y}} \\
&= \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right).
\end{aligned}
$$

Expected Value of a Random Variable

The *expected value*, $\mathrm{E}[X]$ or m_X, also called *expectation*, *average*, or *mean*, of a random variable X is defined as

$$m_X = \mathrm{E}[X] = \sum_x x\, p_X(x)$$

if X is a discrete random variable with PMF $p_X(x)$ and by

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

if X is a continuous random variable with PDF $f_X(x)$.

If $Y = g(X)$, then the expected value of Y can be obtained by finding the PMF $p_Y(y)$ or the PDF $f_Y(y)$ of Y from $p_X(x)$ or $f_X(x)$ and using the above definition, or by finding it directly as $\mathrm{E}[Y] = \sum_x g(x)p_X(x)$ if X is discrete. Indeed,

$$\begin{aligned}
\mathrm{E}[Y] &= \sum_y y p_Y(y) = \sum_y y \sum_{x\in\mathcal{S}_{=y}} p_X(x) \\
&= \sum_y \sum_{x\in\mathcal{S}_{=y}} y p_X(x) = \sum_y \sum_{x\in\mathcal{S}_{=y}} g(x)p_X(x) \\
&= \sum_x g(x)p_X(x),
\end{aligned}$$

as every value x assumed by X belongs to $\mathcal{S}_{=y}$ for some y, namely, $y = g(x)$.

Similarly, if X is continuous, then $\mathrm{E}[Y] = \int_{-\infty}^{\infty} g(x)\, f_X(x)dx$.

Variance, Moments, Moment Generating Function, and Characteristic Function of a Random Variable

The ℓ^{th}-moment of the random variable X, where ℓ is a positive integer, is given by $\mathrm{E}[X^\ell]$. In particular, the first moment is the expectation.

The *variance* of X is $K_X = \mathrm{E}[(X - m_X)^2] = \mathrm{E}[X^2] - m_X^2$.

The *moment generating function* of X is $M_X(s) = \mathrm{E}[e^{sX}]$ where s is in the region of convergence (ROC) of the function[a].

The *characteristic function* of X is obtained by evaluating $M_X(s)$ on the imaginary axis, i.e,. $\Phi_X(\omega) = \mathrm{E}[e^{\jmath\omega X}]$, where $\jmath = \sqrt{-1}$. Since $|e^{\jmath\omega X}| = 1$, the characteristic function always exists.

---

[a]Notice that this is different from G_X defined in (2.53) in the textbook but the same as M_X defined in (3.10).

Provided that the derivatives exist, for every nonnegative integer ℓ,

$$\begin{aligned} \mathrm{E}[X^\ell] &= \frac{d^\ell}{ds^\ell} M_X(s)\bigg|_{s=0}, \\ M_X(s) &= \sum_{\ell \geq 0} \frac{s^\ell}{\ell!} \mathrm{E}[X^\ell]. \end{aligned}$$

If X is discrete,

$$\Phi_X(\omega) = \sum_x p_X(x)\, e^{\jmath \omega x} \quad \text{and} \quad p_X(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_X(\omega) e^{-\jmath \omega x} d\omega.$$

If X is continuous,

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) e^{\jmath \omega x} dx \quad \text{and} \quad f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_X(\omega) e^{-\jmath \omega x} d\omega.$$

Properties of Characteristic Function

i) $\Phi_X(\omega)$ is bounded. In particular, $|\Phi_X(\omega)| \le \Phi_X(0) = 1$.

ii) $\Phi_X(\omega)$ is uniformly continuous over $\mathbb{R}$, i.e., for every $\epsilon > 0$, there exists $\delta > 0$ such that for all real numbers ω_1 and ω_2, if $|\omega_1 - \omega_2| < \delta$, then $|\Phi_X(\omega_1) - \Phi_X(\omega_2)| < \epsilon$.

iii) $\Phi_X^*(\omega) = \Phi_X(-\omega)$.

iv) $\Phi_X(\omega)$ is nonnegative definite (also called positive semi-definite) i.e., for any positive integer n, any real numbers $\omega_1 < \omega_2 < \cdots < \omega_n$, and any (row) complex vector $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{v}[\Phi_X(\omega_l - \omega_m)]_{1\le l,m\le n}\mathbf{v}^{\mathsf{H}} \ge 0$, where H denotes the Hermitian transpose, i.e., the conjugate transpose.

v) $\mathrm{E}[X^\ell] = \frac{1}{\jmath^\ell}\Phi_X^{(\ell)}(0)$, provided that the derivatives exist.

Common Discrete Random Variables

Bernoulli Random Variable A *Bernoulli* random variable X, denoted by Bernoulli(p), is a discrete random variable which assumes two values, 0 and 1, with PMF: $p_X(0) = 1 - p$ and $p_X(1) = p$.
We have

$$\begin{aligned} \mathrm{E}[X] &= p \\ K_X &= p(1-p) \\ M_X(s) &= (1-p) + pe^s \\ \phi_X(\omega) &= (1-p) + pe^{\jmath\omega}. \end{aligned}$$

Binomial Random Variable A *binomial* random variable X, denoted by Binomial(n, p), is a discrete random variable which counts the number of successes in n independent trials where the probability of each success is p. It assumes the values $0, 1, 2, \ldots, n$ with PMF given by

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

We have

$$\begin{aligned} \mathrm{E}[X] &= np \\ K_X &= np(1-p) \\ M_X(s) &= ((1-p) + pe^s)^n \\ \phi_X(\omega) &= ((1-p) + pe^{\jmath\omega})^n. \end{aligned}$$

Poisson Random Variable A *Poisson* random variable X, denoted by Poisson(λ), $\lambda > 0$, is a discrete random variable which assumes the values $0, 1, 2, \ldots$ with PMF given by

$$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

We have

$$\begin{aligned} \mathrm{E}[X] &= \lambda \\ K_X &= \lambda \\ M_X(s) &= e^{\lambda(e^s - 1)} \\ \phi_X(\omega) &= e^{\lambda(e^{\jmath\omega} - 1)}. \end{aligned}$$

Poisson Random Variable as an approximation of a Binomial Random Variable If X is Binomial(n, p) with $p = \lambda/n$ for a fixed λ, then, as $n \to \infty$,

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Indeed,

$$\begin{aligned} p_X(k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

As $n \to \infty$, $n(n-1)\cdots(n-k+1)/n^k \to 1$ and $(1-\frac{\lambda}{n})^{n-k} \to e^{-\lambda}$.

Example 5 The probability that an item is defective is 0.01. Find the probabilities that out of 100 items, exactly three are defective and at most three are defective.

Let X be the number of defective items. Then, X is Binomial$(100, 0.01)$.

$$P(X = 3) = p_X(3) = \binom{100}{3}(0.01)^3(1 - 0.01)^{97} \approx 0.060999.$$

$$\begin{aligned} P(X \leq 3) &= p_X(0) + p_X(1) + p_X(2) + p_X(3) \\ &= \sum_{k=0}^{3} \binom{100}{k}(0.01)^k(0.99)^{100-k} \\ &\approx 0.981626. \end{aligned}$$

With $\lambda = pn = 0.01 \times 100 = 1$, we may approximate X by Poisson(1). Then,

$$P(X = 3) = p_X(3) \approx e^{-1}\frac{1^3}{3!} \approx 0.061313.$$

$$P(X \le 3) = p_X(0) + p_X(1) + p_X(2) + p_X(3) \approx \sum_{k=0}^{3} e^{-1}\frac{1^k}{k!} \approx 0.981012.$$

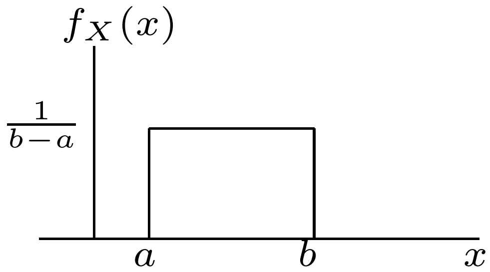
Geometric Random Variable A *geometric* random variable X, denoted by Geometric(p), $0 < p < 1$, is a discrete random variable which counts the number of trials required for the first success in independent trials, where the probability of each success is p. It assumes the values $1, 2, 3, \ldots$ with PMF given by

$$p_X(k) = (1-p)^{k-1}p.$$

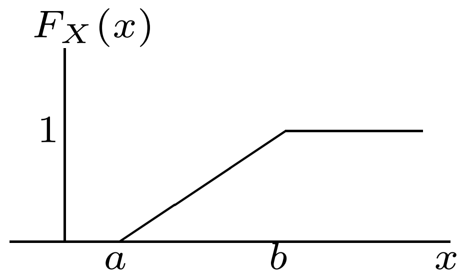
$$\begin{aligned}
\mathrm{E}[X] &= \frac{1}{p} \\
K_X &= \frac{1-p}{p^2} \\
M_X(s) &= \frac{p}{e^{-s}-(1-p)} \quad \text{where real part of } s < -\ln(1-p). \\
\phi_X(\omega) &= \frac{p}{e^{-\jmath\omega}-(1-p)}.
\end{aligned}$$

Common Continuous Random Variables

Uniform Random Variable A *uniform* random variable, X, over the interval $[a, b]$, denoted by $\text{Uniform}(a, b)$, where $a < b$, has PDF and CDF given by



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

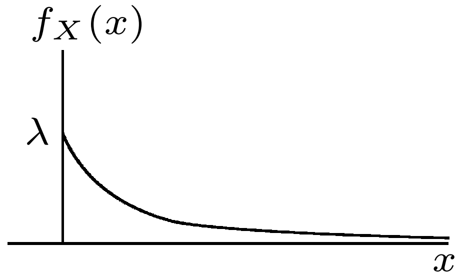


$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b. \end{cases}$$

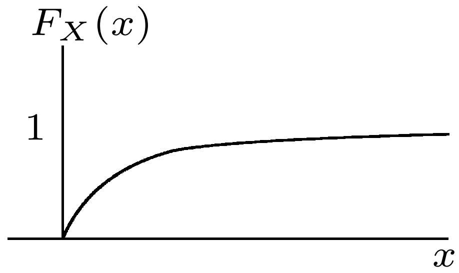
$$
\begin{aligned}
\mathrm{E}[X] &= \int_a^b \frac{x}{b-a}dx = \frac{1}{2}(b+a) \\
\mathrm{E}[X^2] &= \int_a^b \frac{x^2}{b-a}dx = \frac{1}{3}(b^2+ab+a^2) \\
K_X &= \frac{(b-a)^2}{12} \\
M_X(s) &= \begin{cases} \frac{e^{bs}-e^{as}}{s(b-a)} & s \neq 0 \\ 1 & s = 0 \end{cases} \\
\Phi_X(\omega) &= \begin{cases} \frac{e^{\jmath b\omega}-e^{\jmath a\omega}}{\jmath\omega(b-a)} & \omega \neq 0 \\ 1 & \omega = 0. \end{cases}
\end{aligned}
$$

Exponential Random Variable An *exponential* random variable, X, with parameter $\lambda > 0$, denoted by Exponential(λ), has PDF and CDF given by

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



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



$$
\begin{aligned}
\mathrm{E}[X] &= \frac{1}{\lambda} \\
\mathrm{E}[X^2] &= \frac{2}{\lambda^2} \\
K_X &= \frac{1}{\lambda^2} \\
M_X(s) &= \frac{\lambda}{\lambda - s}, \quad \text{where real part of } s < \lambda. \\
\Phi_X(\omega) &= \frac{\lambda}{\lambda - \jmath\omega}.
\end{aligned}
$$

A random variable, X, which assumes only nonnegative values, is *memoryless* if for all $a, b > 0$,

$$P(X > a + b | X > b) = P(X > a),$$

i.e.,

$$P(X > a + b) = P(X > a)\; P(X > b).$$

For example, if the life time of a bulb is memoryless, then the probability that it works for more than $a + b$ years assuming that it worked for more than b years is the same as the probability that it works for more than a years when it was new.

The exponential random variable is memoryless. Indeed,

$$\begin{aligned}
P(X > a + b) &= 1 - F_X(a + b) \\
&= 1 - (1 - e^{-\lambda(a+b)}) \\
&= e^{-\lambda a}\, e^{-\lambda b} \\
&= (1 - F_X(a))\,(1 - F_X(b)) \\
&= P(X > a)\, P(X > b).
\end{aligned}$$

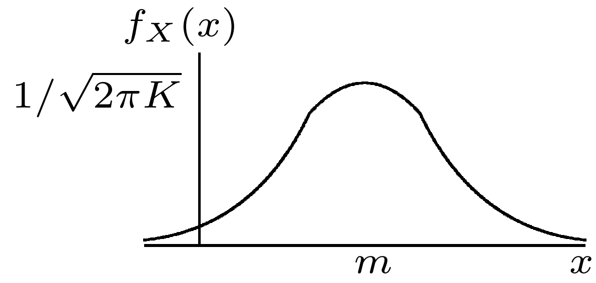
Actually, exponential random variables are the only memoryless continuous random variables. Indeed, suppose X is memoryless and let $h(x) = \ln P(X > x)$. Then, $h(a+b) = h(a) + h(b)$ for $a, b > 0$. We conclude that

$$\frac{h(b)}{b} = \frac{h(a+b) - h(a)}{b}.$$

Pushing b to 0, the right hand side tends to the derivative of $h(x)$ at $x = a$. Since the left hand side does not depend on a, it follows that the derivative of $h(x)$ is constant over all $x \geq 0$. Let $-\lambda$ be this constant. Then, $h'(x) = -\lambda$ and $h(x) = -\lambda x + c$ for some constant c. As $h(x) = \ln P(X > x)$, it follows that $P(X > x) = e^{-\lambda x + c}$ and $F_X(x) = 1 - e^{-\lambda x + c}$. Since $F_X(0) = 0$ as X is a continuous random variable taking only nonnegative values, it follows that $c = 0$. As $F_X(\infty) = 1$, $\lambda > 0$. We conclude that $F_X(x) = 1 - e^{-\lambda x}$ and X is an exponential random variable.

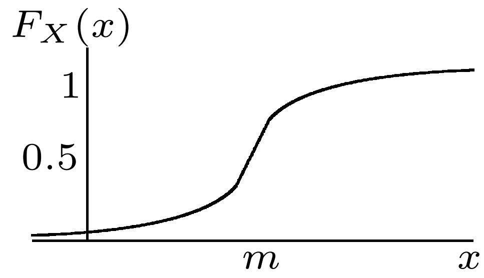
Gaussian Random Variable A *Gaussian*, also called *normal*, random variable, X, denoted by Gaussian(m, K), $K > 0$, is a continuous random variable with PDF and CDF given by

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



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

An important property of Gaussian random variables is that if X is Gaussian(m, K), then $Y = aX + b$, where $a \neq 0$, is Gaussian$(am + b, |a|^2 K)$.



If X is Gaussian(m, K), then

$$\begin{aligned} \mathrm{E}[X] &= m \\ K_X &= K \\ M_X(s) &= e^{ms + \frac{Ks^2}{2}} \\ \Phi_X(\omega) &= e^{\jmath m\omega - \frac{K\omega^2}{2}}. \end{aligned}$$

In particular, m and K are, respectively, the expected value and the variance of Gaussian(m, K).

Let $\varphi(x)$ be the CDF of Gaussian$(0, 1)$, i.e.,

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

Its complement,

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

is called the Q-function. It represents the tail of the Gaussian distribution of Gaussian$(0, 1)$.

Tables for $\varphi(x)$ and $Q(x)$ are widely available.

If X is Gaussian(m, K), then $Y = (X - m)/\sqrt{K}$ is Gaussian$(0, 1)$. We have

$$F_X(x) = \varphi((x - m)/\sqrt{K}) = 1 - Q((x - m)/\sqrt{K}).$$

Bounds on the Q-function We have for $x > 0$,

$$\frac{1}{\sqrt{2\pi}}\frac{x}{x^2+1}e^{-x^2/2} < Q(x) < \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-x^2/2}.$$

In particular,

$$\lim_{x\to\infty} xe^{x^2/2}Q(x) = \frac{1}{\sqrt{2\pi}}.$$

To prove the upper bound, we notice that for $u \geq x > 0$, $\frac{u}{x} \geq 1$. Hence,

$$\begin{aligned} Q(x) &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du \\ &< \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{u}{x} e^{-u^2/2} du \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \end{aligned}$$

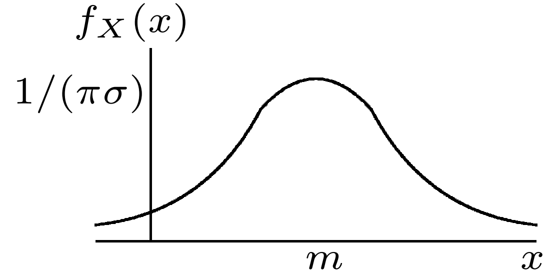
To prove the lower bound, let $g(x) = Q(x) - \frac{1}{\sqrt{2\pi}} \frac{x}{x^2+1} e^{-x^2/2}$. Then,

$$\begin{aligned} g'(x) &= \frac{1}{\sqrt{2\pi}} \left(-e^{-x^2/2} - \frac{d}{dx} \frac{x}{x^2+1} e^{-x^2/2} \right) \\ &= -\frac{2}{\sqrt{2\pi}(x^2+1)^2} e^{-x^2/2} < 0. \end{aligned}$$

Hence, $g(x)$ is a decreasing function. Since $\lim_{x \to \infty} g(x) = 0$, then $g(x) > 0$ for all x.

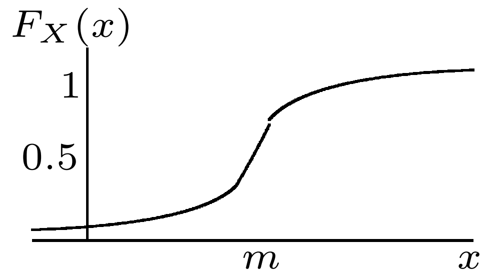
Cauchy Random Variable A *Cauchy* random variable, X, denoted by Cauchy(m, σ^2), $\sigma > 0$, is a continuous random variable with PDF

$$f_X(x) = \frac{1}{\pi\sigma} \frac{1}{1 + \left(\frac{x-m}{\sigma}\right)^2}.$$



Since $f_X(x)$ decays as $1/x^2$, it does not admit any moment. In particular, it has no expectation or variance. We have

$$\begin{aligned} F_X(x) &= \frac{1}{\pi} \arctan\left(\frac{x-m}{\sigma}\right) + \frac{1}{2} \\ \Phi_X(\omega) &= e^{\jmath\omega m - \sigma|\omega|}. \end{aligned}$$



$M_X(s)$ exists only if s is purely imaginary in which case it reduces to $\Phi_X(\omega)$.

Joint Distribution of Random Variables

Consider the discrete random variables X and Y. The *joint probability mass function* (joint PMF) of X and Y is the probability $P(X = x, Y = y)$ that X takes the value x and Y takes the value y for any possible values x and y assumed by X and Y, respectively. This probability is denoted by $p_{X,Y}(x, y)$.

Consider the continuous random variables X and Y. The *joint cumulative distribution function* (joint CDF) is the probability $P(X \leq x, Y \leq y)$ for real x and y. This probability is denoted by $F_{X,Y}(x, y)$.

We say that X and Y are *jointly continuous* if there exists a function $f_{X,Y}(x, y)$ such that for all $x, y \in \mathbb{R}$,

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du.$$

The function $f_{X,Y}(x, y)$ is called the *joint probability density function* (joint PDF) of X and Y. If $F_{X,Y}(x, y)$ is differentiable, then

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

Properties of the Joint Cumulative Distribution Function

i) $F_{X,Y}(x,-\infty) = F_{X,Y}(-\infty,y) = 0$, $F_{X,Y}(\infty,\infty) = 1$.

ii) $F_{X,Y}(x,\infty) = F_X(x)$ and $F_{X,Y}(\infty,y) = F_Y(y)$.

iii) $F_{X,Y}(x,y)$ is nondecreasing in the two variables x and y, i.e., $F_{X,Y}(x_1,y_1) \leq F_{X,Y}(x_2,y_2)$ for $x_1 \leq x_2$ and $y_1 \leq y_2$.

We have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

$F_X(x)$ and $F_Y(y)$ are called *marginal CDF*s and $f_X(x)$ and $f_Y(y)$ are called *marginal PDF*s of X and Y, respectively.

Independent Random Variables

The discrete random variables X and Y are *independent* if

$$p_{X,Y}(x, y) = p_X(x)\, p_Y(y)$$

for all values x and y assumed by X and Y, respectively.

The continuous random variables X and Y are *independent* if

$$F_{X,Y}(x, y) = F_X(x)\, F_Y(y),$$

or equivalently,

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

for all real numbers x and y.

The *conditional probability mass function* of the discrete random variable X given the discrete random variable Y is

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

provided that $p_Y(y) \neq 0$.

X and Y are independent if and only if $p_{X|Y}(x|y) = p_X(x)$.

The *conditional cumulative distribution function* of the continuous random variable X given the continuous random variable Y is

$$F_{X|Y}(x|y) = \frac{F_{X,Y}(x,y)}{F_Y(y)}$$

provided that $F_Y(y) \neq 0$.

The *conditional probability density function* of the continuous random variable X given the continuous random variable Y is

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

provided that $f_Y(y) \neq 0$.

X and Y are independent if and only if $F_{X|Y}(x|y) = F_X(x)$, or equivalently, $f_{X|Y}(x|y) = f_X(x)$.

Bayes' Rule for PDFs

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx}.$$

The *covariance* of X and Y is

$$K_{X,Y} = E[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

We have

$$|K_{X,Y}| \leq \sqrt{K_X K_Y}.$$

Indeed, for any a,

$$\begin{aligned} 0 \leq \mathrm{E}[(a(X - \mathrm{E}[X]) - (Y - \mathrm{E}[Y]))^2] \\ = a^2 K_X - 2aK_{X,Y} + K_Y. \end{aligned}$$

In particular, as a polynomial in a, the above polynomial has no distinct real roots. Hence, its discriminant, $4K_{X,Y}^2 - 4K_X K_Y$, is nonpositive and the result follows.

The ratio $K_{X,Y}/\sqrt{K_X K_Y}$ is called the *correlation coefficient* of X and Y, denoted by $\rho_{X,Y}$. We have $|\rho_{X,Y}| \leq 1$.

If X and Y are independent, then

$$K_{X,Y} = 0.$$

We say that X and Y are *uncorrelated* if $K_{X,Y} = 0$.

Notice that X and Y are uncorrelated if and only if $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$ for any functions $g(x) = ax + b$ and $h(y) = cy + d$, where a, b, c, and d are constants.

However, X and Y are independent if and only if $\mathrm{E}[g(X)h(Y)] = \mathrm{E}[g(X)]\mathrm{E}[h(Y)]$ for any functions $g(x)$ and $h(y)$, i.e., not necessarily linear.[a]

---

[a] See Section 2.9.2 for proofs of both statements.

Conditional Expectation Let X and Y be random variables. We define the *conditional expectation* of X given that $Y = y$ to be

$$\mathrm{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

or

$$\mathrm{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx,$$

depending on whether X and Y are discrete or continuous.

Define the random variable $\mathrm{E}[X|Y]$ as follows. Given an outcome ω that corresponds to a value y, the random variable $\mathrm{E}[X|Y]$ assumes the value $\mathrm{E}[X|Y=y]$. The expected value of this random variable is

$$\begin{aligned}
\mathrm{E}[\mathrm{E}[X|Y]] &= \int_{-\infty}^{\infty} \mathrm{E}[X|Y=y] f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X,Y}(x,y) dy dx \\
&= \int_{-\infty}^{\infty} x f_X(x) dx = \mathrm{E}[X]
\end{aligned}$$

in case X and Y are continuous. The result $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$ holds also for discrete random variables. It allows for the iterative computation of the mean of a random variable.

Example 6 A drunk man reaches a street intersection. If he turns right, he will reach his destination in an average of 10 minutes. If he drives straight, he will reach it in an average of 20 minutes. If he turns left, he will return to the intersection without reaching his destination in an average of 30 minutes. He will then again choose one of the three options until he reaches his destination. What is the expected time it takes the man to reach his destination if each time he reaches the intersection he chooses one of the three options with probability 1/3?

Let X be the number of minutes it takes the man to reach his destination. Let Y be the option he chooses the first time he reaches the intersection which assumes the values 1, 2, and 3 corresponding to right, straight, and left, respectively. We have $\mathrm{E}[X|Y=1]=10$, $\mathrm{E}[X|Y=2]=20$, and $\mathrm{E}[X|Y=3]=30+\mathrm{E}[X]$. The random variable $\mathrm{E}[X|Y]$ assumes the values 10, 20, and $30+\mathrm{E}[X]$ with equal probabilities. Hence,

$$\mathrm{E}[X]=\mathrm{E}[\mathrm{E}[X|Y]]=\frac{1}{3}(10+20+30+\mathrm{E}[X]).$$

Solving this equation gives $\mathrm{E}[X]=30$.

Joint Characteristic Function

The *joint characteristic function* of the continuous random variables X and Y is given by

$$\Phi_{X,Y}(\mu,\nu) = \mathrm{E}[e^{\jmath(\mu X+\nu Y)}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)e^{\jmath(\mu x+\nu y)}dxdy.$$

Notice that X and Y are independent if and only if

$$\Phi_{X,Y}(\mu,\nu) = \Phi_X(\mu)\Phi_Y(\nu).$$

Sum of Independent Random Variables Let X and Y be independent random variables and $Z = X + Y$. Then

$$\Phi_Z(\omega) = \mathrm{E}[e^{\jmath\omega(X+Y)}] = \mathrm{E}[e^{\jmath\omega X}]\ \mathrm{E}[e^{\jmath\omega Y}] = \Phi_X(\omega)\ \Phi_Y(\omega).$$

If X and Y are discrete, then Z is discrete with PMF given by

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x).$$

This sum is called the *convolution sum* of the PMFs of X and Y.

If X and Y are continuous, then Z is continuous with PDF given by

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx.$$

This integral is called the *convolution integral* of the PDFs of X and Y.

We have $K_{X+Y} = K_X + K_Y$ if X and Y are independent.

Notice that $m_{X+Y} = m_X + m_Y$ regardless if X and Y are independent or not.

Transformation of Random Variables Let X_1 and X_2 be jointly continuous random variables, $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be functions of X_1 and X_2, which are also jointly continuous random variables. Let $(x_{1,i}, x_{2,i}) = (h_{1,i}(y_1, y_2), h_{2,i}(y_1, y_2))$, for $i = 1, 2, \ldots$, be the solutions of the equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$. Assume that g_1 and g_2 have continuous partial derivatives and

$$\det \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix} \neq 0$$

for $(x_{1,i}, x_{2,i}) = (h_{1,i}(y_1, y_2), h_{2,i}(y_1, y_2))$, $i = 1, 2, \ldots$. This determinant, denoted by $J(x_1, x_2)$, is called the *Jacobian* of the transformation defined by g_1 and g_2.

Then,

$$f_{Y_1,Y_2}(y_1, y_2) = \sum_i \frac{f_{X_1,X_2}(x_{1,i}, x_{2,i})}{|J(x_{1,i}, x_{2,i})|}\Bigg|_{x_{1,i}=h_{1,i}(y_1,y_2),x_{2,i}=h_{2,i}(y_1,y_2)}.$$

Example 7 Let $R = \sqrt{X^2 + Y^2}$ and $\Theta = \arctan(Y/X)$. Then, for $r > 0$ and $0 \le \theta < 2\pi$, $r = \sqrt{x^2 + y^2}$ and $\theta = \arctan(y/x)$ have the solution $x = r\cos(\theta)$ and $y = r\sin(\theta)$. Also,

$$J(x, y) = \det \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y/x^2}{1+(y/x)^2} & \frac{1/x}{1+(y/x)^2} \end{pmatrix} = \frac{1}{\sqrt{x^2 + y^2}}.$$

Hence,

$$\begin{aligned} f_{R,\Theta}(r, \theta) &= \left. \frac{f_{X,Y}(x, y)}{1/\sqrt{x^2 + y^2}} \right|_{x=r\cos(\theta), y=r\sin(\theta)} \\ &= r\, f_{X,Y}(r\cos(\theta), r\sin(\theta)). \end{aligned}$$

If X and Y are independent Gaussian$(0,1)$ random variables, then

$$\begin{aligned} f_{R,\Theta}(r,\theta) &= r\,\frac{1}{\sqrt{2\pi}}e^{-(r\cos(\theta))^2/2} \times \frac{1}{\sqrt{2\pi}}e^{-(r\sin(\theta))^2/2} \\ &= \frac{1}{2\pi} r e^{-r^2/2} \end{aligned}$$

for $r > 0$ and $0 \leq \theta < 2\pi$. In particular, R and Θ are independent, Θ is uniform, and R has a PDF given by $f_R(r) = re^{-r^2/2}$, $r > 0$.

A random variable, R, with such a PDF is called a *Rayleigh* random variable.

Random Vectors

Let $X_1, X_2, \ldots, X_n$ be random variables. Then, $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, is a (row) *random vector*.

If $X_1, X_2, \ldots, X_n$ are discrete, then $\mathbf{X}$ is a discrete random vector with joint PMF

$$p_{\mathbf{X}}(\mathbf{x}) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $x_1, x_2, \ldots x_n$ are values assumed by $X_1, X_2, \ldots, X_n$, respectively.

If $X_1, X_2, \ldots, X_n$ are continuous, then $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is a continuous random vector with joint CDF given by

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

The joint PDF of $\mathbf{X}$, denoted by $f_{\mathbf{X}}(\mathbf{x})$, is given by

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{X_1, X_2, \ldots, X_n}(u_1, u_2, \ldots, u_n) du_1 du_2 \cdots du_n.$$

If $F_{\mathbf{X}}(\mathbf{x})$ is differentiable, then

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\mathbf{X}}(x_1, x_2, \ldots, x_n).$$

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. Then

$$\mathbf{m_X} = (m_1, m_2, \ldots, m_n)$$

is called the *mean vector* of $\mathbf{X}$ and

$$\mathbf{K_X} = \begin{pmatrix} K_{X_1} & K_{X_1,X_2} & \ldots & K_{X_1,X_n} \\ K_{X_2,X_1} & K_{X_2} & \ldots & K_{X_2,X_n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{X_n,X_1} & K_{X_n,X_2} & \ldots & K_{X_n} \end{pmatrix}$$

is called the *covariance matrix* of $\mathbf{X}$. It is a symmetric $n \times n$ matrix in which the (i, j) entry is K_{X_i,X_j}. It is nonnegative definite (also called positive semi-definite) and, in particular, its determinant, $\det(\mathbf{K_X})$, is nonnegative.

First, we show that $\mathbf{K_X}$ is nonnegative definite, i.e., for any vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ of real numbers, we have $\mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T} \geq 0$. Indeed,

$$
\begin{aligned}
\mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T} &= \mathbf{a}\mathrm{E}[(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\mathsf{T}(\mathbf{X} - \mathrm{E}[\mathbf{X}])]\mathbf{a}^\mathsf{T} \\
&= \mathrm{E}[\mathbf{a}(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\mathsf{T}(\mathbf{X} - \mathrm{E}[\mathbf{X}])\mathbf{a}^\mathsf{T}] \\
&= \mathrm{E}[(\mathbf{a}(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\mathsf{T})(\mathbf{a}(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\mathsf{T})^\mathsf{T}] \\
&= \mathrm{E}[(\mathbf{a}(\mathbf{X} - \mathrm{E}[\mathbf{X}])^\mathsf{T})^2] \\
&= \mathrm{E}[(a_1(X_1 - \mathrm{E}[X_1]) + \cdots + a_n(X_n - \mathrm{E}[X_n]))^2] \geq 0,
\end{aligned}
$$

as it is the expectation of a square.

Next, we show that $\det(\mathbf{K_X}) \geq 0$. Assume to get a contradiction that $\det(\mathbf{K_X}) < 0$. Consider the function $f(t) = \det(t\mathbf{I} + (1-t)\mathbf{K_X})$, where $\mathbf{I}$ is the $n \times n$ identity matrix. Being a polynomial in t, this function is continuous. We have $f(0) = \det(\mathbf{K_X}) < 0$ and $f(1) = \det(\mathbf{I}) = 1$. By the intermediate value theorem, there exists t_0, $0 < t_0 < 1$, such that $f(t_0) = 0$, i.e., $t_0\mathbf{I} + (1-t_0)\mathbf{K_X}$ is singular. Hence, there is a nonzero vector $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ such that $\mathbf{a}(t_0\mathbf{I} + (1-t_0)\mathbf{K_X})$ is the all-zeros vector and, therefore, $\mathbf{a}(t_0\mathbf{I} + (1-t_0\mathbf{K_X}))\mathbf{a}^\mathsf{T} = 0$. However,

$$\mathbf{a}(t_0\mathbf{I} + (1-t_0)\mathbf{K_X})\mathbf{a}^\mathsf{T} = t_0\mathbf{a}\mathbf{a}^\mathsf{T} + (1-t_0)\mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T} > 0,$$

as $0 < t_0 < 1$, $\mathbf{a}\mathbf{a}^\mathsf{T} = a_1^2 + a_2^2 + \cdots + a_n^2 > 0$, and $\mathbf{K_X}$ is positive semi-definite. This contradiction proves that $\det(\mathbf{K_X}) \geq 0$.

The components of $\mathbf{X} = (X_2, X_2, \ldots, X_n)$ are *independent* if and only if for every $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we have

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1}(x_1)p_{X_2}(x_2)\cdots p_{X_n}(x_n)$$

or

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

depending on whether $\mathbf{X}$ is discrete or continuous, respectively.

If $\mathbf{Y}$ is a continuous random vector of length n which is a function of a continuous random vector $\mathbf{X}$ of length n, then the PDF of $\mathbf{Y}$ may be obtained from that of $\mathbf{X}$ by using the Jacobian which is the determinant of an $n \times n$ matrix of partial derivatives.

Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. Let $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ and $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$.

Then, the moment generating function of $\mathbf{X}$ is

$$M_{\mathbf{X}}(\mathbf{s}) = \mathrm{E}[e^{\mathbf{X}\mathbf{s}^\mathsf{T}}] = \mathrm{E}[e^{s_1 X_1 + s_2 X_2 + \cdots + s_n X_n}]$$

and its characteristic function is

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = \mathrm{E}[e^{\jmath \mathbf{X}\boldsymbol{\omega}^\mathsf{T}}] = \mathrm{E}[e^{\jmath(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)}].$$

The random variables $X_1, X_2, \ldots, X_n$ are independent if and only if

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = \prod_{i=1}^{n} \Phi_{X_i}(\omega_i).$$

Notice that if $X = X_1 + X_2 + \cdots + X_n$, then

$$\mathrm{E}[X] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]$$

regardless if $X_1, X_2, \ldots, X_n$ are independent or not.

If $X_1, X_2, \ldots, X_n$ are independent, then we also have

$$K_X = K_{X_1} + K_{X_2} + \cdots + K_{X_n}.$$

Gaussian Random Vectors The vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of random variables is called a *Gaussian random vector* if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{K_X})}} e^{-\frac{1}{2}\left((\mathbf{x}-\mathbf{m_X})\mathbf{K_X}^{-1}(\mathbf{x}-\mathbf{m_X})^{\mathsf{T}}\right)},$$

where $\det(\mathbf{K_X}) > 0$ and T denotes transpose. In this case, we say that $X_1, X_2, \ldots, X_n$ are *jointly Gaussian*.

For $n = 2$,

$$f_{X_1,X_2}(x_1, x_2) =$$

$$\frac{1}{2\pi\sqrt{K_{X_1}K_{X_2}}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-m_1)^2}{K_{X_1}} - \frac{2\rho(x_1-m_1)(x_2-m_2)}{\sqrt{K_{X_1}K_{X_2}}} + \frac{(x_2-m_2)^2}{K_{X_2}}\right)}$$

where

$$\begin{aligned}
\mathbf{K}_{X_1,X_2} &= \begin{pmatrix} K_{X_1} & \rho\sqrt{K_{X_1}K_{X_2}} \\ \rho\sqrt{K_{X_1}K_{X_2}} & K_{X_2} \end{pmatrix}, \\
\mathbf{K}^{-1}_{X_1,X_2} &= \frac{1}{1-\rho^2} \begin{pmatrix} 1/K_{X_1} & -\rho/\sqrt{K_{X_1}K_{X_2}} \\ -\rho/\sqrt{K_{X_1}K_{X_2}} & 1/K_{X_2} \end{pmatrix},
\end{aligned}$$

and ρ is the correlation coefficient of X_1 and X_2.

The components $X_1, X_2, \ldots, X_n$ of the Gaussian random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are independent if and only if they are uncorrelated, i.e., $K_{X_i,X_j} = 0$ for all $i \neq j$. In this case, $\mathbf{K_X}$ is a diagonal matrix and

$$
\begin{aligned}
& f_{\mathbf{X}}(\mathbf{x}) \\
= \quad & \frac{1}{\sqrt{(2\pi)^n K_{X_1} K_{X_2} \cdots K_{X_n}}} e^{-\frac{1}{2}\left(\frac{(x_1-m_1)^2}{K_{X_1}} + \frac{(x_2-m_2)^2}{K_{X_2}} + \cdots + \frac{(x_n-m_n)^2}{K_{X_n}}\right)} \\
= \quad & \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi K_{X_i}}} e^{-\frac{(x_i-m_i)^2}{2K_{X_i}}},
\end{aligned}
$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

Characteristic Function of a Gaussian Random Vector Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a Gaussian random vector with mean vector $\mathbf{m}_X$ and covariance matrix $\mathbf{K}_{\mathbf{X}}$. Then, its characteristic function is

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = e^{\jmath \mathbf{m}_{\mathbf{X}} \boldsymbol{\omega}^{\mathsf{T}} - \frac{\boldsymbol{\omega} \mathbf{K}_{\mathbf{X}} \boldsymbol{\omega}^{\mathsf{T}}}{2}},$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_n)$.

The Sum of Jointly Gaussian Random Variables is Gaussian Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a Gaussian random vector with mean vector $\mathbf{m}_X$ and covariance matrix $\mathbf{K_X}$. Let

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n = \mathbf{X}\mathbf{a}^T,$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ is a vector of n real numbers. Then, Y is Gaussian$(\mathbf{m_X}\mathbf{a}^\mathsf{T}, \mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T})$. Indeed,

$$\begin{aligned}
\Phi_Y(\omega) &= \mathrm{E}[e^{\jmath\omega Y}] = \mathrm{E}[e^{\jmath\omega(a_1X_1+a_2X_2+\cdots+a_nX_n)}] \\
&= \mathrm{E}[e^{\jmath((\omega a_1)X_1+(\omega a_2)X_2+\cdots+(\omega a_n)X_n)}] \\
&= \Phi_{\mathbf{X}}(\omega a_1, \omega a_2, \ldots, \omega a_n) \\
&= \Phi_{\mathbf{X}}(\omega\mathbf{a}) \\
&= e^{\jmath\mathbf{m_X}\mathbf{a}^\mathsf{T}\omega - \frac{\mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T}\omega^2}{2}},
\end{aligned}$$

which is the characteristic function of Gaussian$(\mathbf{m_X}\mathbf{a}^\mathsf{T}, \mathbf{a}\mathbf{K_X}\mathbf{a}^\mathsf{T})$.

Affine Transformation of a Gaussian Random Vector Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a Gaussian random vector with mean vector $\mathbf{m}_X$ and covariance matrix $\mathbf{K}_\mathbf{X}$. Let $\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{b}$, where $\mathbf{A}$ is an invertible $n \times n$ real matrix and $\mathbf{b}$ is a real vector of length n. Then, $\mathbf{Y}$ is an affine transformation of $\mathbf{X}$.

We will show that $\mathbf{Y}$ is a Gaussian random vector with mean vector $\mathbf{m}_\mathbf{Y} = \mathbf{m}_\mathbf{X}\mathbf{A} + \mathbf{b}$ and covariance matrix $\mathbf{K}_\mathbf{Y} = \mathbf{A}^\top \mathbf{K}_\mathbf{X} \mathbf{A}$.

Indeed,

$$\begin{aligned}
\Phi_{\mathbf{Y}}(\boldsymbol{\omega}) &= \mathrm{E}\left[e^{\jmath \mathbf{Y}\boldsymbol{\omega}^\top}\right] = \mathrm{E}\left[e^{\jmath (\mathbf{X}\mathbf{A}+\mathbf{b})\boldsymbol{\omega}^\top}\right] \\
&= e^{\jmath \mathbf{b}\boldsymbol{\omega}^\top}\mathrm{E}\left[e^{\jmath \mathbf{X}\mathbf{A}\boldsymbol{\omega}^\top}\right] \\
&= e^{\jmath \mathbf{b}\boldsymbol{\omega}^\top}\mathrm{E}\left[e^{\jmath \mathbf{X}(\boldsymbol{\omega}\mathbf{A}^\top)^\top}\right] \\
&= e^{\jmath \mathbf{b}\boldsymbol{\omega}^\top}\Phi_{\mathbf{X}}(\boldsymbol{\omega}\mathbf{A}^\top) \\
&= e^{\jmath \mathbf{b}\boldsymbol{\omega}^\top}\ e^{\jmath \mathbf{m}_{\mathbf{X}}(\boldsymbol{\omega}\mathbf{A}^\top)^\top - \frac{(\boldsymbol{\omega}\mathbf{A}^\top)\mathbf{K}_{\mathbf{X}}(\boldsymbol{\omega}\mathbf{A}^\top)^\top}{2}} \\
&= e^{\jmath (\mathbf{m}_{\mathbf{X}}\mathbf{A}+\mathbf{b})\boldsymbol{\omega}^\top - \frac{\boldsymbol{\omega}(\mathbf{A}^\top\mathbf{K}_{\mathbf{X}}\mathbf{A})\boldsymbol{\omega}^\top}{2}},
\end{aligned}$$

which is the characteristic function of a Gaussian random vector with mean vector $\mathbf{m}_{\mathbf{Y}} = \mathbf{m}_{\mathbf{X}}\mathbf{A} + \mathbf{b}$ and covariance matrix $\mathbf{K}_{\mathbf{Y}} = \mathbf{A}^\top\mathbf{K}_{\mathbf{X}}\mathbf{A}$.

Examples of Gaussian Random Variables that are not Jointly Gaussian:

Example 1 Let X be the Gaussian random variable Gaussian$(0, 1)$. Let A be a random variable, independent of X, that assumes the values -1 and 1 with equal probability $1/2$. Let $Y = AX$. We will show that Y is Gaussian$(0, 1)$. Indeed, we have $P(Y \leq y|A = 1) = P(X \leq y)$ and

$$P(Y \leq y|A = -1) = P(-X \leq y) = P(X \geq -y) = P(X \leq y)$$

by symmetry of the PDF of X. Hence,

$$P(Y \leq y) = \frac{1}{2}P(Y \leq y|A = 1) + \frac{1}{2}P(Y \leq y|A = -1) = P(X \leq y).$$

This proves that Y is Gaussian$(0, 1)$. Notice that X and Y are not jointly Gaussian. Indeed, if they were, then $Z = X + Y$ is Gaussian. However,

$$Z = \begin{cases} 2X, & \text{if } A = 1 \\ 0, & \text{if } A = -1. \end{cases}$$

Hence, $P(Z = 0) = 1/2$ which contradicts the fact that the probability that a Gaussian random variable assumes any particular value is 0.

Example 2 Let X and Y be jointly continuous random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} e^{-(x^2+y^2)/2} & xy \geq 0 \\ 0 & xy < 0. \end{cases}$$

Since $f_{X,Y}(x,y)$ is not the joint PDF of jointly Gaussian random variables, X and Y are not jointly Gaussian. We check that X is Gaussian. For $x \geq 0$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^{\infty} \frac{1}{\pi} e^{-(x^2+y^2)/2} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \times 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$

On the other hand, for $x < 0$,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{0} \frac{1}{\pi} e^{-(x^2+y^2)/2} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \times 2 \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$

We conclude that $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for all x and X is Gaussian. Similarly, we can check that Y is Gaussian. However, X and Y are not jointly Gaussian. Notice also that the sum $Z = X + Y$ is not Gaussian. Indeed, we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)dx.$$

Hence, $f_Z(0) = 0$ since $f_{X,Y}(x,y) = 0$ if $xy < 0$. Clearly, Z is not Gaussian since a Gaussian random variable has a strictly positive PDF.