REVIEW OF PROBABILITY
DEFINITION

Experiments are a set S of elements of outcomes. Some subsets are called events. The space S is the certain event, and the empty set is the impossible event. If an event $\{\alpha_i\}$ contains a single element α it is called an elementary event. Two events are mutually exclusive if they do not have common events.

We assign to each event A a number $P(A)$ called the probability of the event A. The probability satisfies the following three conditions:

(i) Non-Negativity: $P(A) \geq 0$.

(ii) Unit Summation: $P(S) = 1$.

(iii) Mutual Exclusion: If $A \cap B = 0$, then $P(A + B) = P(A) + P(B)$.

Given a set event M, with $P(M) > 0$, we define the conditional probability of A given M as

$$P(A|M) = \frac{P(AM)}{P(M)}.$$

If A and M are mutually exclusive, then we have that

$$P(A|M) = 0.$$

If $A \subset M$, then $P(A|M) \geq P(A)$. If $M \subset A$, then $P(A|M) = 1$. These numbers are also probabilities.

Given n mutually exclusive events $A_1, A_2, \ldots, A_N$, their sum equals S, i.e.,

$$\sum_i A_i = 1,\ A_i A_j = 0,\ i \neq j.$$

Then the probability of an arbitrary event B can be obtained as

$$P(B) = P(B|M)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_N)P(A_N).$$

Instead of working on the element space, one can work on the real line by defining a transformation from the event space to the real line. Suppose an experiment ξ specified by a space S, and a field $\mathcal{F}$ of subsets in S called events, a probability P is assigned to each of these events. To every outcome of this experiment we assign a real number $x(\xi)$. Therefore we have defined a function $x : S \leftarrow \mathcal{R}$.

This function is called a random variable (r.v.) if:

(i) The set $\{X \leq x\}$ is an event for any real number x.

(ii) the probability of the event $\{x = +\infty\}$ and $\{x = -\infty\}$ are zero:
$P(X = +\infty) = P(X = -\infty) = 0.$

REVIEW OF PROBABILITY
DISTRIBUTION FUNCTION

Given real numbers x, the set $\{X \leq x\}$ consists of all outcomes ξ such that $x(\xi) \leq x$ is an event. Its probability $P(X \leq x)$ is a number dependent on x (a function of x) denoted as $F_X(x)$ and it is called the distribution function. By definition, $F_X(x) = P(X \leq x)$, for $x \in [-\infty, \infty]$.

There are a few important properties of $F_X(x)$:

(i) Left Nullity: $F(-\infty) = 0.$

(ii) Right Unity: $F(\infty) = 1.$

(iii) Non-Decreasing: $F(x_1) \leq F(x_2)$ if $x_1 < x_2$.

(iv) Right Continuity: $F(X^+) = F(X)$.

The density function of a random variable x is

$$f(x) = \frac{dF_X(x)}{dx}.$$

There may be difficulties in computing this function. Suppose $f(x)$ is a continuous function of x, but it may not be differentiable everywhere. If the number of non-differentiable points are countable, then x is a continuous random variable. The density function is defined in all differentiable points in the domain.

There are two key properties of the density function:

(i) Non-Negativity: $f(x) \geq 0$.

(ii) Unit Summability: $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

The meaning of $f(x)$ is really the probability in the delta interval, i.e., If the distributions is of a staircase type, the random variable is called discrete. If p_i is the jump of $F(x)$ at x_i,

$$P(\{X = x_i\}) = p_i = F(x_i) - F(x_i^-), i = 1, 2, ...$$

and

$$\sum_i p_i = F(\infty) - F(-\infty) = 1, F(x) = \sum_i P(X = x_i)$$

The probability density for discrete random variable is normally called the probability mass function because the impulses $p_i\delta(x - x_i)$ can be considered as point masses p_i placed at x_i. There may be random variables of the mixed type.

REVIEW OF PROBABILITY

EXAMPLES OF DENSITY FUNCTIONS

Gaussian or Normal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $F(x) = \int_{-\infty}^{x} f(y)\,dy = \frac{1}{2} + \exp\left(\frac{x-\mu}{\sigma}\right)$
Poisson (single event)	$f(x) = \exp^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \delta(x-k)$ $P(X=k) = \frac{\exp^{\alpha^k}}{k!}, k = 0, 1, \ldots\ \alpha > 0$
Binomial (dual events)	$P(X=k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}, p+q=1$ $f(x) = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k} \delta(x-k)$
Uniform	$f(x) = \begin{cases} 1 & x_1 \leq x \leq x_2 \\ 0 & \text{otherwise} \end{cases}$
Gamma	$f(x) = Ax^b \exp^{-cx} U(x)$

REVIEW OF PROBABILITY

CONDITIONAL DISTRIBUTIONS AND DENSITIES

Let $P\{M\} > 0$ where M is an event set. The condition distribution function $F_X(x|M)$ is defined as the conditional probability of the event $\{X \leq x\}$

$$F_X(x|M) = P(X < x|M) = \frac{P(X < x, M)}{P(M)}$$

where $\{X \leq x, M\}$ is the set $\{x(\xi) \leq x$ and $\xi \in M\}$. This means we are restricting the event set to M, so all definitions of probability apply to $F_X(x|M)$. Likewise for densities:

$$f(x|M) = \frac{dF_X(x|M)}{dx}$$

REVIEW OF PROBABILITY
FUNCTION OF ONE RANDOM VARIABLE

Let x be a random variable (r.v.) and $y = g(x) = g(x(\xi))$ is a function of a r.v., so y is a r.v.. Let us define I_y as the set of x for $g(x) \leq y$. We define $F_Y(y) = P(Y \leq y) = P(x \in I_y)$.

$$\text{Example: } g(x) = \begin{cases} x + c & x \geq 0 \\ x - c & x < 0 \end{cases}$$

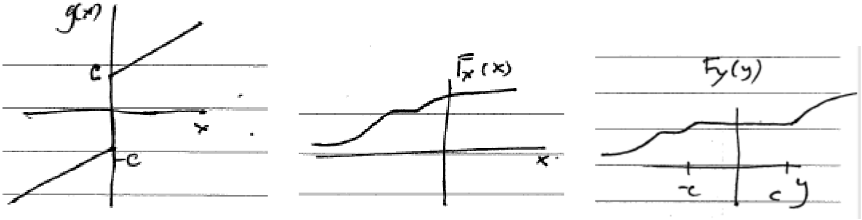


Figure: Left: $g(x)$ function; center: distribution function of r.v. X; and right: distribution

△ △ △

One of the most important function of a random variable. By definition, the expected value of a r.v. X is

$$E[X] = \int_{\infty}^{+\infty} xf(x)\,dx \text{ (continuous r.v.)}$$

$$E[X] = \sum_n x_n P(X = x_n) = \sum_n x_n p_n \text{ (discrete r.v.)}$$

Is $\eta = E[X]$ a r.v.? NO! y is deterministic because we are integrating over x using all the knowledge we have about x ($f(x)$).

Sometimes in math books we find

$$E[X] = \int xdF(x)$$

▽ ▽ ▽

The variance of a r.v. X is defined as

$$E[(X-\mu)^2] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx \text{ (moment of inertia)}$$

where μ is the mean of the r.v. X, that is, $\mu = E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$. The variable σ is called the **standard deviation**. For discrete variables,

$$\sigma^2 = E[(X-\mu)^2] = \sum_n (x_n-\mu)^2 P(X=x_n), \text{ and}$$

$$\mu = E[X] = \frac{1}{N}\sum_{i=1}^{N} x_i P(X=x_i)$$

Note that $\nabla^2 = E[X^2] - E^2[X]$.

A more complete specification of the statistics of a r.v. X is acquired by the moments:

$$m_k = E[X^k] = \int_{-\infty}^{\infty} x^k f(x)\,dx$$

So $m_1 = \mu$. The central moments are defined as

$$\mu_k = E[(X-\mu)^k] = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx\,(\mu_2 = \sigma^2)$$

We can also write

$$\mu_k = \sum_{n=0}^{k} \left(\begin{array}{c} k \\ n \end{array} \right) (-1)^n \mu^n m_{n-k}$$

Define $\Phi_X(u) = E[\exp^{jux}] = \int_{-\infty}^{\infty} f(x) \exp^{jux} dx$, which is the Fourier Transform (FT) of $f(x)$ with a reversal in sign. For discrete r.v. $\Phi_X(u) = \sum_k \exp^{jux_k} P(X = x_k)$. So using the FT pair

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi(u) \exp^{-jux} du$$

By the moment theorem

$$\text{If } E[X^k] = m_k$$

$$\text{then } \frac{d^k\Phi(0)}{du^k} = \int m_k$$

There are many ideas that can be defined for cases of two or more random variables. Let X, Y be two r.v. with distribution functions $P(X \leq x) = F_X(x)$ and $P(Y \leq y) = F_Y(y)$, respectively. The joint distribution is defined as

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

Likewise, the joint pdf is defined as

$$f(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

Joints are two dimensional functions for scalars x and y.

$$\begin{cases} f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy \\ f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx \end{cases} .$$

Assume that $M = \{X \leq x\}$ and $P(M) \neq 0$ then

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

Likewise

$$\mathbf{f}_Y(y|X \leq x) = \frac{\frac{\partial F_{XY}(x,y)}{\partial y}}{F_X(x)} = \frac{\int_{-\infty}^{x} f_{XY}(\xi, y)\, d\xi}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(\xi, y)\, d\xi\, dy}$$

But the event M can be defined in any way one wants. If $M = \{X \in I\}$ where I is a single point x, then $M = \{X = x\}$ with $f(x) \neq 0$.

$$F_Y(y|X = x) = \lim_{\triangle x \to 0} F_Y(y|x < X < x + \triangle x) \text{ or}$$

$$F_Y(y|X = x) = \frac{\int_{-\infty}^{y} f_{XY}(x, p)\, dp}{\ldots}$$

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

Again, consider the set $M = \{X \in I\} = \{X = x\}$ and $I = \{x\}\}$

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

where $E[Y|X = x]$ is the center of mass of the r.v. Y given $X = x$.

Now we will define $E[Y|X]$ which is a r.v. (instead of a number as before) defined as: For a given experimental outcome ξ the r.v. x taks the value $x(\xi)$. The corresponding value of $E[Y|X] = E[Y|X = x(\xi)]$. So it is a function of the r.v. x,

$$E[Y|X = x] = \int^{\infty} yf(y|x)\, dy.$$

Two events A and B are independent if $P(AB) = P(A)P(B)$. The r.v. X is independent of Y if

$$P(X \leq z, Y \leq y) = P(X \leq x)P(Y \leq y)$$

and it follows that

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

and

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

So the joint distribution becomes the product of the marginal distributions and can be directly computed from them, which is a huge simplicity factor.

REVIEW OF PROBABILITY

JOINTLY NORMAL RANDOM VARIABLES

$$f_{XY}(x,y) = A\exp^{-\frac{ax^2+bxy+cy^2+dx+ey}{B}}$$

where

$$B = \frac{1}{2(1-r^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(x-\mu_2)^2}{\sigma_2^2} - 2r\frac{(x-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2}\right]$$

The marginal distributions are Gaussian and $f_{XY}(x,y) = f_X(x)f_Y(y)$ if $r = 0$ and r is called the correlation coefficient.

REVIEW OF PROBABILITY
FUNCTIONS OF TWO R.V.

Let $z = g(x, y)$, that is, $z(\xi) = g(x(\xi), y(\xi))$. Define $\{Z \leq z\} = \{(X, Y) \in D_Z\}$ then

$$F_Z(z) = P(Z \leq z) = P(X, Y \in D_Z) = \int\int_{D_Z} f_{XY}(x, y)\, dx dy$$

Example:

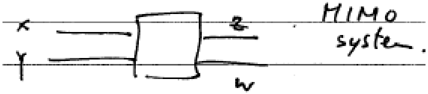
$$z = x + y$$

$$F_Z(z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx dy \text{ and}$$

$$f(z) = \int_{-\infty}^{\infty} f_{XY}(z - y, y)\, dy$$

REVIEW OF PROBABILITY
TWO FUNCTIONS OF TWO R.V.

Let X and Y be two r.v., now consider $z = g(x, y)$ and $w = h(x, y)$. This is a system called Multiple Input Multiple Output (MIMO) system. Z and W have a joint distribution $F_{ZW}(z, w)$ and density $f_{ZW}(z, w)$.



How can we find them if we know that functions g and h are rational and we also know $f_X(x)$ and $f_Y(y)$? By definition,

$$F_{ZW}(z, w) = \int\int_{D_{ZW}} f_{XY}(x, y)\, dxdy$$

with $D_{ZW} = \begin{cases} g(x, y) \leq z \\ h(x, y) \leq w \end{cases}$

EXPECTED VALUE OF TWO R.V.

Let $z = g(x, y)$. The expected value of the r.v. Z is

$$E[Z] = \int_{-\infty}^{\infty} z f_Z(z)\, dz$$

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

REVIEW OF PROBABILITY
CONDITIONAL EXPECTED VALUE

The conditional expected value $E[g(x, y)|X = x]$ can be determined from $f_{XY}((x, y)|X = x)$, but it is a 2D input, so we use lists

$$f_{XY}((x, y)|x_1 \leq X \leq x_2) = \frac{f_{XY}(x, y)}{F_X(x_2) - F_X(x_1)}$$

So

$$E[g(X, Y)|x_1 \leq X \leq x_2] = \int_{-\infty}^{\infty} \int_{x_1}^{x_2} \frac{g(x, y)F_{XY}(x, y)}{F_X(x_2) - F_X(x_1)} dxdy$$

Now taking $\triangle x = x_2 - x_1 \to 0$

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

Note that $g(X, Y)$ and $g(x, Y)$ represent two different things. The first is a r.v. function of 2

$$
\begin{aligned}
m_{kr} &= E[X^k Y^r] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^k y^r f_{XY}(x, y)\, dx dy
\end{aligned}
$$

Example: For $k = 1$ and $r = 1$, m_{11} is also known as the correlation between the r.v. X and the r.v. Y.

$$m_{11} = E[XY] = R_{XY}.$$

$$\mu_{kr} = E[(X - \mu_X)^k (Y - \mu_Y)^r]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)^k (y - \mu_Y)^r f_{XY}(x, y)\, dxdy$$

Example: For $k = 1$ and $r = 1$, μ_{11} is also known as the covariance between the r.v. X and the r.v. Y.

$$\mu_{11} = E[(X - \mu_X)(Y - \mu_Y)] = cov(X, Y)$$

The ratio

$$r = \frac{E\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sqrt{E\left[(X - \mu_X)^2\right] E\left[(Y - \mu_Y)^2\right]}}$$

is the correlation coefficient and $|r| \leq 1$. When $r = 0$, X and Y are uncorrelated r.v. and it means that $E[XY] = E[X]E[Y]$. The r.v. X and Y are orthogonal if $E[XY] = 0$. The r.v. X

$$\Phi_{XY}(u_1, u_2) = E\left[\exp^{j(u_1 x + u_2 y)}\right]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp^{j(u_1 x + u_2 y)} f_{XY}(x, y)\, dx dy$$

The marginal characteristic function is

$$\Phi_X(u) = \Phi_{XY}(u_1, 0)$$

If X and Y are independent:

$$\Phi_{XY}(u_1, u_2) = \Phi_X(u_1)\Phi_Y(u_2)$$

Example: If X and Y are independent and their sum is Poisson distributed, then X and Y are Poisson distributed.

Let us approximate a r.v. X by a constant (simplest case). What should this constant be? Answer: it depends on the criterion we use to rate success! Let us use Mean Square Error (or MSE) which is defined as:

$$E\left[(x-a)^2\right] = \int_{-\infty}^{\infty} (x-a)^2 f_X(x)\,dx$$

In words, MSE minimizes the variance of the errors $(x_i - a)$. So we need to find a constant a that minimizes the MSE criterion

$$\min_a E\left[(X-a)^2\right]$$

We maintain that $a = E[X] = \mu_X$ meets the goal. The derivation is as follows:

$$E[(X-a)^2] = E[X^2 - 2aX + a^2] = E[X^2] - 2aE[X] + a^2$$

Now we want to estimate Y by an unknown and potential nonlinear function $g(X)$, such that

$$E[(Y - g(X))^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y - g(x))^2 f_{XY}(x, y)\,dxdy$$

The function $g(x)$ that minimizes

$$\min_{g(x)} E[(y - g(x))^2]$$

is the conditional expected value of Y given X, that is

$$g(X) = E[Y|X]$$

In fact, $f_{XY}(x, y) = f_Y(Y|X)f_X(x)$ so

△ △ △

Problem: evaluate $E[Y|X]$. The linear MSE provides a simple solution which is better than a constant! Consider $y = ax + b$

$$\min_{a,b} E[(y-(ax+b))^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-(ax+b))^2 f_{XY}(x,y)\,dxdy$$

The solution is given by

$$a = \frac{r\sigma_Y}{\sigma_X}$$
$$b = E[Y] - aE[X]$$

and the minimum error is

$$e_{min} = \sigma_Y^2(1-r^2)$$

(r is the c.c. between X and Y)

▽ ▽ ▽