Probability Primer[1]

- **Sample space and Events**
  - Suppose that an experiment with an uncertain outcome is performed (e.g., rolling a die).
  - While the outcome of the experiment is not known in advance, the set of all possible outcomes is known. This set is the sample space, Ω.
  - For example, when rolling a die $\Omega = \{1, 2, 3, 4, 5, 6\}$. When tossing a coin, $\Omega = \{H, T\}$. When measuring life time of a machine (years), $\Omega = \{1, 2, 3, \ldots\}$.
  - A subset $E \subset \Omega$ is known as an event.
  - For example, when rolling a die, $E = \{1\}$ is the event that one appears. The subset $E = \{1, 3, 5\}$ is the event that an odd number appears.

- **Probability of an event**
  - If an experiment is repeated for a number of times which is large enough, the fraction of time that event E occurs is the probability that event E occurs, $P\{E\}$.
  - E.g., when rolling a fair die, $P\{1\} = 1/6$, and $P\{1, 3, 5\} = 3/6 = 1/2$. When tossing a fair coin, $P\{H\} = P\{T\} = 1/2$.

[1] Compiled from S. M. Ross, *Introduction to Probability Models*, 6th Edition, Academic Press, 1997.

- **Axioms of probability**

(1) For $E \subseteq \Omega$, $0 \le P\{E\} \le 1$;

(2) $P\{\Omega\} = 1$;

(3) For events $E_1, E_2, \ldots, E_i, \ldots$, with $E_i \subset \Omega$, $E_i \cap E_j = \varnothing$, for all i and j, $P\left\{\bigcup_{i=1}^{\infty} E_i\right\} = \sum_{i=1}^{\infty} P\{E_i\}$.

- **Implications**
  - The axioms of probability imply the following results:
    - For E and $F \subset \Omega$,

      $P\{E \text{ “or” } F\} = P\{E \cup F\} = P\{E\} + P\{F\} - P\{E \cap F\}$;[2]
    - If E and F are mutually exclusive (i.e., $E \cap F = \varnothing$), then

      $P\{E \cup F\} = P\{E\} + P\{F\}$;
    - For $E \subset \Omega$, let E^c be the complement of E (i.e., $E \cup E^c = \Omega$),

      $P\{E^c\} = 1 - P\{E\}$;
    - $P\{\varnothing\} = 0$.

- **Conditional probability**
  - The probability that event E occurs given that event F has already occurred is

$$P\{E \mid F\} = \frac{P\{E \cap F\}}{P\{F\}} .$$

  - E.g., when rolling two fair dice, suppose the first die is 3, what is the probability the sum of the two dice is 7?

---

[2] $P\{E \cap F\} = P\{E \text{ “and” } F\}$.

- Let E be the event that the sum of the two dice is 7, $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, and F be the event that the first die is 3, $F = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$. Then,

$$P\{E \mid F\} = \frac{P\{E \cap F\}}{P\{F\}} = \frac{P\{(3,4)\}}{P\{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}}$$

$$= \frac{1/36}{6/36} = \frac{1}{6}.$$

- **Independent events**
  - For E and $F \subset \Omega$, $P\{E \cap F\} = P\{E|F\}P\{F\}$.
  - Two events are independent if an only if $P\{E \cap F\} = P\{E\}P\{F\}$. That is, $P\{E|F\} = P\{E\}$.

- **Random variables**
  - Consider a function that assigns real numbers to events (outcomes) in Ω. Such real-valued function is a *random variable*.
  - E.g., when rolling two fair dice, define X as the sum of the two dice. Then, X is a random variable with $P\{X = 2\} = P\{(1,1)\} = 1/36$, $P\{X = 3\} = P\{(1, 2), (2, 1)\} = 2/36 = 1/18$, etc.
  - E.g., the salvage value of a machine, S, is \$1,500 if the market goes up (with probability 0.4) and \$1,000 if the market goes down (with probability 0.6). Then, S is a random variable with $P\{S = 1500\} = 0.4$ and $P\{S = 1000\} = 0.6$.

- If a random variable can take on a limited number of values. Then, this is a *discrete random variable*. E.g., the random variable X representing the sum of two dice.
- If the random variable can take on an uncountable number of values. Then, this is a continuous random variable. E.g., the random variable H representing height of an AUB student.
- If X is a discrete random variable, the function $f_X(x) = P\{X = x\}$ is the *probability mass function* (pmf) of X.
- The function $F_X(x) = \mathrm{P}\{X \le x\} = \sum_{x_i \le x} f_X(x_i)$ is the *cumulative distribution function* (CDF) of X.
- CDF is sometimes simply referred to as *distribution function*.
- E.g., for the random variable S representing salvage value of a machine above,

$$f_S(s) = \begin{cases} 0.6 & \text{if } s = 1000 \\ 0.4 & \text{if } s = 1500 \\ 0 & \text{othewise} \end{cases} \quad , \quad F_S(s) = \begin{cases} 0 & \text{if } \ s < 1000 \\ 0.6 & \text{if } \ 1000 \le s < 1500 \\ 1 & \text{if } \ s \ge 1500 \end{cases} .$$

- For a continuous random variable, the CDF is defined based on a function, $f_X(x)$,the *probability density function* (pdf),

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

- **Independent Random variables**
    - Two random variables X and Y are said to be independent if

$$P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\} = F_X(x)F_Y(y) .$$

- **Expectation of a random variable**
  - The *expectation* of a discrete random variable X is

$$E[X] = \sum_{x_i} x_i P\{X = x_i\} = \sum_{x_i} x_i f_X(x_i) .$$

  - The expectation of a continuous random variable X is

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

  - The expectation of a random variable is the value obtained if the underlying experience is repeated for a number of times which is large enough and the resulting values are averaged.
  - The expectation is "linear." That is, for two random variables X and Y, $E[aX + bY] = aE[X] + bE[Y]$.

- **Measures of variability**
  - The *variance* of a discrete random variable X is

$$\mathrm{Var}[X] = E[(X - E[X])^2] = \sum_{x_i} x_i^2 f_X(x_i) - \left(E[X]\right)^2 .$$

  - The variance of a continuous random variable X is

$$\mathrm{Var}[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} x^2 f_X(x) - \left(E[X]\right)^2 .$$

  - The *standard deviation* of a random variable X is
    $\sigma_X = \sqrt{\mathrm{Var}[X]}$.
  - The *coefficient of variation* of a random variable X is CV[X] $= \sigma_X / E[X]$.

- The variance (standard deviation) measures the spread of the random variable around the expectation.
- The coefficient of variation is useful when comparing the variability of different alternatives.
- Note that Var[aX] =a^2 Var[X], for any real number a and random variable a .

- **Joint distribution**
  - The joint distribution function of two random variables is

$$F_{X,Y}(x,y) = P\{X \le x, Y \le y\}.$$

  - If X and Y are discrete random variables then,

$$F_{X,Y}(x,y) = \sum_{i \le x, j \le y} P\{X = i, Y = j\} = \sum_{i \le x, j \le y} f_{X,Y}(i,j),$$

where $f_{X,Y}(.)$ is the joint pmf of X and Y.

  - If X and Y are continuous random variables then,

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

where $f_{X,Y}(.)$ is the joint pdf of X and Y.

Fact. $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ *if and only if (iff) X and Y are independent***.**

- **Covariance**
  - The covariance measures the dependence of two random variables. For two random variables X and Y,

$$\begin{aligned}\sigma_{XY} &= \mathrm{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y],\end{aligned}$$

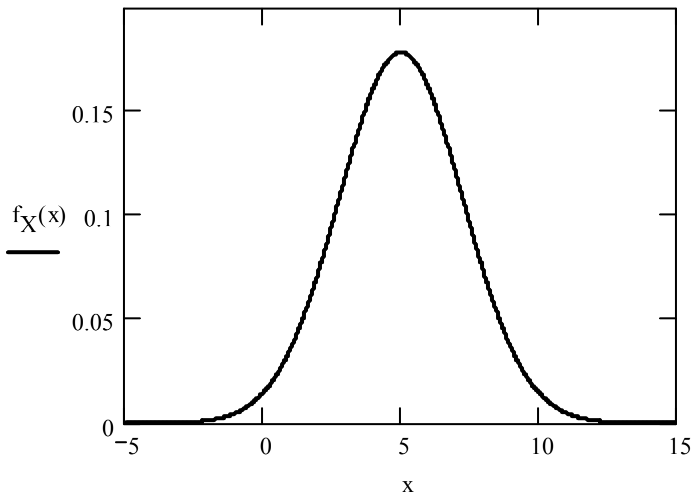
    where,

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

  - If $\sigma_{XY} > 0$ (<0), X and Y are said to be positively (negatively) correlated.
  - $\sigma_{XY} = 0$ iff X and Y are independent.
  - The coefficient of correlation is defined as $\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}$.
  - Note that $\rho_{XY} \leq 1$.
  - Note that $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + 2\mathrm{Cov}[X, Y] + \mathrm{Var}[Y]$.
  - If X and Y are independent, $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

- **The Normal Random Variable**
  - We say that a random variable X is a normal rv with parameters μ and $\sigma > 0$ if it has the following pdf:

$$f_X(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma}, \ x \in (-\infty, \infty).$$



- Note that $f_X(x)$ defines a pdf. With a change of variable $z = (x - \mu)/\sigma$ and using the fact that $\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}$,

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1 .$$

- It can be also shown by doing the appropriate integration that $E[X] = \mu$ and $\mathrm{Var}[X] = \sigma$.
- The normal rv is a good model for quantities that can be seen as sums or averages of a large number of rv's.
- The cdf of X, $F_X(x) = \int_{-\infty}^{x} f_X(t) dt$, has no closed-form.

Fact. *If X is a normal rv, then $Z = (X - \mu)/\sigma$ is a "standard normal r.v." with parameters 0 and* 1.

- The cdf of X, $F_X(x)$, is then evaluates through the cdf of Z, $F_Z(z)$, which is often tabulated, as

$$P\{X < x\} = P\{Z < \frac{x-\mu}{\sigma}\} \Rightarrow F_X(x) = F_Z\left(\frac{x-\mu}{\sigma}\right) .$$