Total probability formula and Bayes Theorem

Recall for two events A, B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$\Rightarrow P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Suppose now the sample space S can be divided into sets: $B_1 \cup B_2 \cup \ldots \cup B_K = S$; B_j mutually exclusive.

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_K)$$

$$\Rightarrow P(A) = \sum_{i=1}^{K} P(A \cap B_i) = \sum_{i=1}^{K} P(B_i)\, P(A|B_i)$$

"Total probability formula"

Bayes' Rule

Assume $S = B_1 \cup B_2 \cup \dots \cup B_k$; B_j disjoint $1 \le j \le k$

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)}$$

$$= \frac{P(B_j)\,P(A \mid B_j)}{\sum_{i=1}^{k} P(B_i)\,P(A \mid B_i)}$$

Ex) Suppose good and defective microchips are sorted in 3 boxes

5 defective 20 good	10 defective 20 good	5 defective 35 good
Box 1	Box 2	Box 3

Choose a box at random, then select a microchip from a box. Calculate the probability that microchip came from box 1, given that it is defective.

(That is, suppose we do this in dark and draw a faulty microchip, find probability it came from Box 1)

B_i : chip came from box i

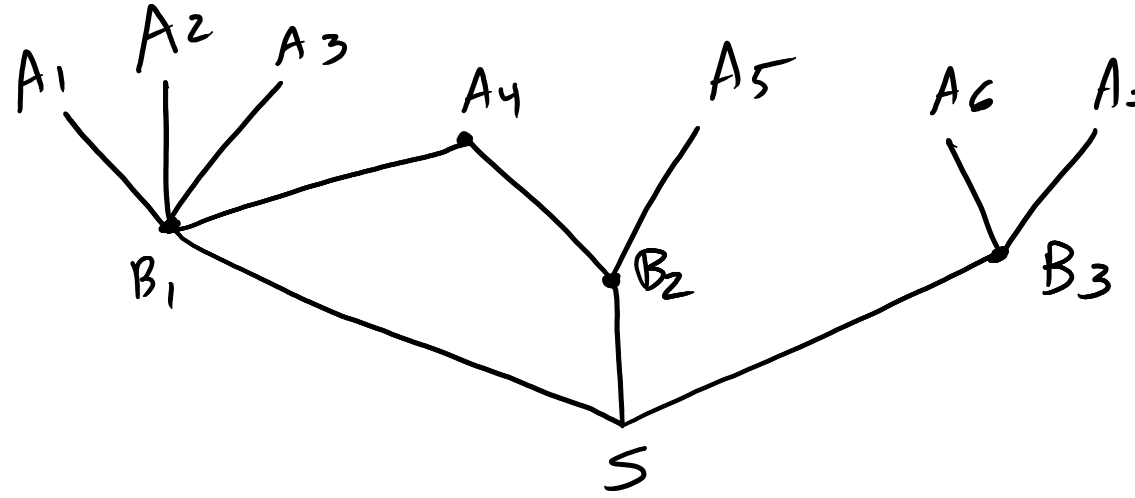
A : chip is defective

$$P(B_1 | A) = \frac{P(B_1)\,P(A|B_1)}{\sum_{i=1}^{3} P(B_i)\,P(A|B_i)}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{5}{25}\right)}{\left(\frac{1}{3}\right)\left(\frac{5}{25}\right)+\left(\frac{1}{3}\right)\left(\frac{10}{35}\right)+\left(\frac{1}{3}\right)\left(\frac{5}{40}\right)}$$

$$= \frac{56}{171} \approx 0.327$$

Ex) A man starts at the map shown below at point S and follows it to point B_1, B_2, or B_3. From that point, he chooses a new path at random and follows it to one of the points A_i, $i = 1, 2, \ldots, 7$.



What is the probability that the man arrives at point A_4?

Here, we use the law of total probability:

$$P(A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

$$P(A_4) = P(B_1)P(A_4|B_1) + P(B_2)P(A_4|B_2) + P(B_3)P(A_4|B_3)$$

$$= \left(\frac{1}{3}\right)\left(\frac{1}{4}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{2}\right) + \left(\frac{1}{3}\right)(0) = \frac{1}{4}$$

Suppose now the man arrives at point A_4, but it is <u>not</u> known which route he took. The probability that he passed through a particular point $B_1, B_2,$ or B_3 can be computed from Bayes' Rule:

$$P(B_1 | A_4) = \frac{(1/3)(1/4)}{(1/3)(1/4) + (1/3)(1/2) + (1/3)(0)} = \frac{1}{3}$$

$$\left(= \frac{P(B_1)\,P(A_4 | B_1)}{\sum_{i=1}^{3} P(B_i)\,P(A_4 | B_i)} \right)$$

Discrete random variables review

If the set of all possible values of a random variable X is a <u>countable</u> set $\{x_1, x_2, \dots\}$ (notice, not necessarily finite) then X is called a discrete random variable.

The function $f(x) = P[X = x]$ is called the discrete probability density function.

$$0 \leq f(x_i) \leq 1$$

$$\sum_i f(x_i) = 1$$

The cumulative distribution function (CDF) is defined for any real x as:

$$F(x) = P[X \leq x]$$

Ex) Roll a four sided (tetrahedral) die twice. The die has numbers $\{1,2,3,4\}$ affixed to each of its four sides. Each is equally likely to come up.

Score = max of two numbers which occur.

Let X be r.v. recording the score.

$P(X=0) = 0$

$P(x=1) = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16} \quad \{(1,1)\}$

$P(x=2) : \{(1,2), (2,1), (2,2)\}; \; P(x=2) = \frac{3}{16}$

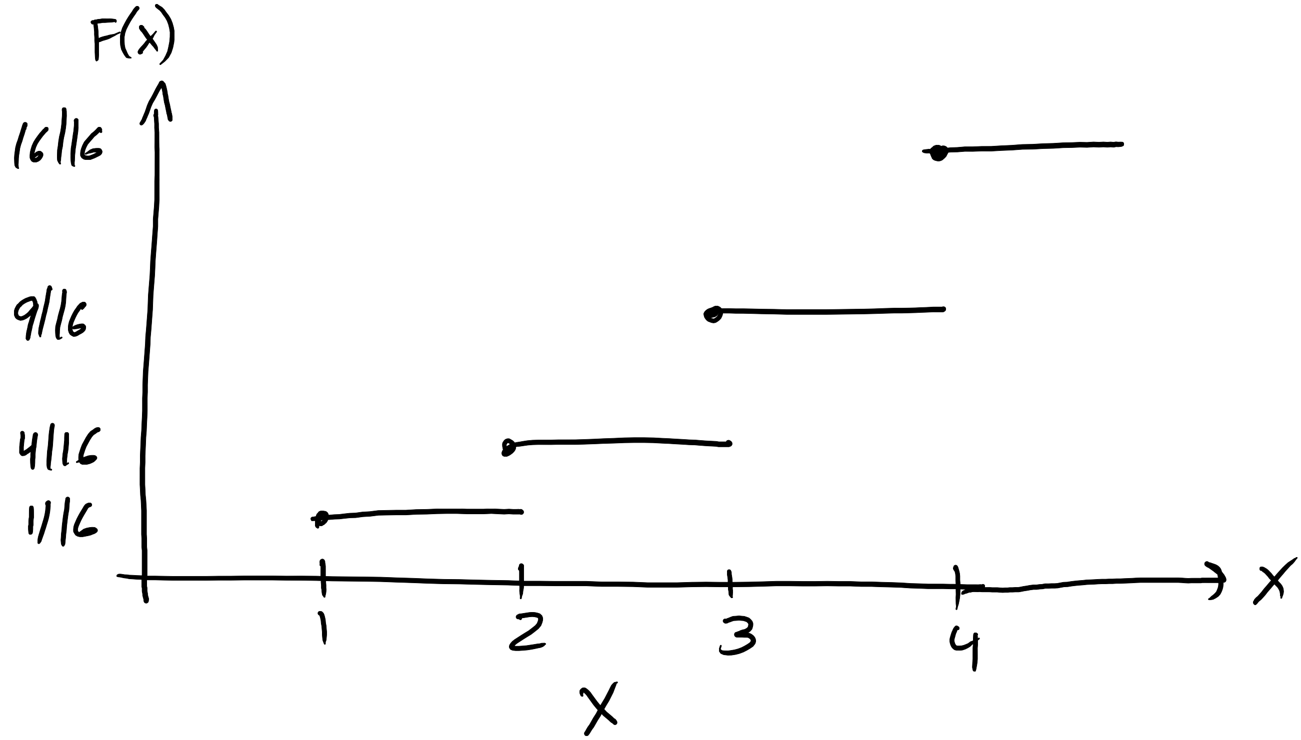
$P(x=3) : \{(3,1), (1,3), (2,3), (3,2), (3,3)\}$

$$P(x=3) = \frac{5}{16}$$

$x=4 : \{\underset{1}{(4,1)}, \underset{2}{(4,2)}, \underset{3}{(4,3)}, \underset{4}{(1,4)}, \underset{5}{(2,4)}, \underset{6}{(3,4)}, \underset{7}{(4,4)}\}$

$P(x=4)=\frac{7}{16}$

$F(x)=P(X\leq x)$. If we plot the CDF of max of two rolls of a four sided die it will look like a non-decreasing step function



$F(0.5)=P(x\leq 0.5)=0$

$F(1.5)=P(x\leq 1.5)=P(x\leq 1)=\frac{1}{16}$

$F(3.5)=P(x\leq 3)=P(x=1)+P(x=2)+P(x=3)=\frac{9}{16}$

$F(10)=P(x\leq 10)=P(x\leq 4)=1$

Expected value and variance:

$$E[X] = \sum_x x f(x) = \sum_x x P(X=x)$$

Can be considered as a measure of the "center" of the associated probability distribution.

$$\mu_X = E[X] = 1 \cdot \frac{1}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{5}{16} + 4 \cdot \frac{7}{16}$$

$$= \frac{25}{8} = 3.125$$

$$\sigma_X^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x)$$

$$= \sum_x (x-\mu)^2 P(X=x)$$ <u>variance</u>

$$= \left(1-\frac{25}{8}\right)^2 \frac{1}{16} + \left(2-\frac{25}{8}\right)^2 \frac{3}{16} + \left(3-\frac{25}{8}\right)^2 \frac{5}{16}$$

$$+ \left(4-\frac{25}{8}\right)^2 \frac{7}{16} \approx 0.86$$

$$\Rightarrow \sigma_X = \sqrt{\sigma_X^2} \approx 0.93$$ <u>std deviation</u>

Law of large numbers and simulation

Suppose the two dice game is played many times. Let the (scores) be the values of the random variables:

$$X_1, X_2, \dots, X_n$$

Law of large numbers

If $X_1, \dots, X_n$ is a random sample from a distribution with finite mean μ and variance σ^2, then the sequence of sample means converges in probability to μ.

"sample mean", itself a random variable

Let $\overline{X}_n = \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n}\sum X_i$

$$E[\overline{X}_n] = \frac{1}{n}\left(E[X_1] + \dots + E[X_n]\right) = \frac{1}{n}(n\mu)$$

$$= \mu$$

for indep r.v.'s $X_1, \dots, X_n$

On the other hand,

$$Var(\bar{X}_n) = Var\left[\frac{1}{n}\sum X_i\right]$$

Now, for independent random variables X and Y:

$$Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y)$$

$$\Rightarrow Var(\bar{X}_n) = \underbrace{\frac{1}{n^2}\sigma_X^2 + \ldots + \frac{1}{n^2}\sigma_X^2}_{n \text{ terms}} = \frac{1}{n}\sigma_X^2$$

$$= \frac{1}{n} Var(X)$$

Chebyshev's Inequality

Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any real number $k>0$:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Notice: only the case $K>1$ is useful.
When $K \le 1$, the right hand side $\frac{1}{K^2} \ge 1$ and the inequality trivially holds since all probabilities are between 0 and 1.

Apply Chebyshev's inequality on $\overline{X}_n$:

$$P(|\overline{X}_n - \mu| \ge K\sigma_{\overline{X}_n}) \le \frac{1}{K^2}$$

Let $K\sigma_{\overline{X}_n} = \varepsilon \Rightarrow K^2 = \frac{\varepsilon^2}{\sigma_{\overline{X}_n}^2}$

$$\Rightarrow K^2 = \frac{\varepsilon^2}{\frac{1}{n}\sigma_X^2} = \frac{n\varepsilon^2}{\sigma_X^2}$$

$$\Rightarrow P(|\overline{X}_n - \mu| \ge \varepsilon) \le \frac{\sigma_X^2}{n\varepsilon^2}$$

Thus,

$$P(|\overline{X}_n - \mu| < \varepsilon) = 1 - P(|\overline{X}_n - \mu| \ge \varepsilon)$$

So that:

$$P(|\bar{X}_n - \mu| < \varepsilon) \geq 1 - \frac{\sigma_X^2}{n\varepsilon^2}$$

Notice that as $n \to \infty$, no matter how small $\varepsilon > 0$ is chosen,

$$P(|\bar{X}_n - \mu| < \varepsilon) \to 1$$

This is what is meant by "convergence in probability" in the Law of Large Numbers.

$$\bar{X}_n \xrightarrow{P} \mu \quad \text{as } n \to \infty$$

(Let's do a simulation in R).

Simulation 1: Rolling two 4 sided dice game as above.

Notice that as the number of trials N increases, the value of the sample mean $\bar{X}_N$ gets very close to $E[X]$, where $\bar{X} \sim dist(\mu, \sigma)$ we have derived.