# CS146 Assignment 1

## 1. Rules of probability theory

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

This statement is **correct** because we can derive it from the definition of conditional probability. One note is that when `P(B) = 0`, this equation is still correct because `P(A ∩ B)` would then also be `0` because event `B` never occurs.

---

$$P(A) = P(A|B)P(B) + P(A|\neg B)P(\neg B) = P(A,B) + P(A,\neg B) = P(A, B \cup \neg B) = P(A, \mathbb{U}) = P(A)$$

This statement is **correct**. When we marginalize over `B` and `¬B`, we have covered the whole universe of events.

---

$$P(A) = P(A|B)P(B) + P(A|C)P(C) + P(A|D) + P(D)$$

This statement is **incorrect** because events `B`, `C`, and `D` do not necessarily cover the whole universe of events. Thus, when we marginalize `A` over those, we will not get the full probability of `A`.

---

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

This statement is **correct** because it is equivalent to the definition of conditional probability. We go through the seconds equality step by referring back to the first equation. Third equality shows the commutative feature of joint probability because they are the exact same event.

---

$$P(A|B)P(B) = P(B|A)P(A) = P(A,B) = P(B,A)$$

This statement is **correct**. Using the proof in the first equation, we can see that the left side is equal to `P(A, B)` and the right side is equal to `P(B, A)`. These two, in turn, are equal to each other because

## 2. Logarithms and probability distributions

$$f_n(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$

Now, we need to grab the natural logarithm (base `e`) of this equation:

$$ln(f_n) = ln(\frac{1}{\sqrt{2\pi\sigma^2}}) + ln(exp[-\frac{(x-\mu)^2}{2\sigma^2}]) = ln(1) - ln(\sqrt{2\pi\sigma^2}) - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}ln(2\pi\sigma^2) = ln(\frac{1}{\sigma}) - \frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}ln(2\pi)$$

---

$$f_\gamma(x|\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

$$ln(f_\gamma) = ln(\beta^\alpha) + ln(x^{\alpha-1}) + ln(e^{-\beta x}) - ln(\Gamma(\alpha)) = \alpha ln(\beta) + (\alpha-1)ln(x) - \beta x - ln(\Gamma(\alpha))$$

---

$$f_\beta(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)x^{\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}$$

$$ln(f_\beta) = ln(\Gamma(\alpha+\beta)) + ln(x^{\alpha-1}) + ln((1-x)^{\beta-1}) - ln(\Gamma(\alpha)) - ln(\Gamma(\beta)) = (\alpha-1)ln(x) + (\beta-1)ln(x) + ln(\Gamma(\alpha+\beta)) - ln(\Gamma(\alpha)) - ln(\Gamma(\beta))$$

## 3. Normal distribution

$$f(x) = x^3 + 2x + 1,\ x \sim N(\mu,\sigma)$$

To find the expected value of `x` as a random variable, we would sum up the products of event value (`x`) and probability (`p(x)`) of that event. In a continuous case, this would become:

$$\int_{-\inf}^{+\inf} xp(x)dx = \int_{-\inf}^{+\inf} \frac{x}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x-\mu)^2}{2\sigma^2}]dx$$

Now, in case of `f(x)`, its probabilities depend on the probabilities of `x`, thus, we can re-use the `p(x)` for corresponding `f(x)`:

$$E(f) = \int_{-\inf}^{+\inf} \frac{x^3+2x+1}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x-\mu)^2}{2\sigma^2}]dx$$

---

Bringing in the graph (Fig. 1) of the given function in Desmos, we can see that this function is monotonically growing (has only one root), and `f(x) = 1` when `x = 0`. It is apparent that we can prove this by checking that the derivative (`3x² + 2`) is always positive.

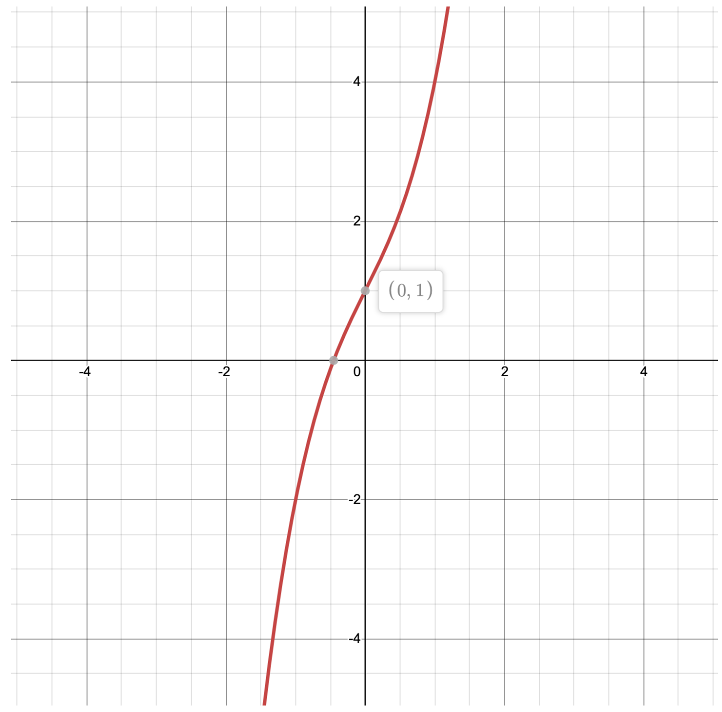


Fig. 1: Graph of f(x) = x³ + 2x + 1. Produced in desmos.com.

Given the information about function behavior, we can paraphrase the question in terms of an inequality for `x`:

$$P(f(x) > 1) = P(x > 0) = \int_0^{+\inf} \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x-\mu)^2}{2\sigma^2}]dx$$

For this example, let's pick `mu = -1` and `sigma = 2` and calculate (with WolframAlpha) the probability:

$$P(f(x) > 1 | \mu = -1, \sigma = 2) = \int_0^{+\inf} \frac{1}{\sqrt{8\pi}} exp[-\frac{(x+1)^2}{8}]dx \approx 0.309$$

---

Let is cross-check with a Python script:

```
import numpy as np

mu, sigma = -1, 2
samples = np.random.normal(mu, sigma, 1000000) # generate 1 million samples
f_vals = samples**3 + 2*samples + 1
np.sum(f_vals > 1) / f_vals.shape[0]

>>> 0.309071
```

Looks like our mathematical derivation was correct!

## 4. Marginal and conditional probabilities

$$P(young) = 0.252$$

---

$$P(\neg young) = 1 - P(young) = 0.748$$

---

$$P(unemployed) = P(unemployed|young)P(young) + P(unemployed|\neg young)P(\neg young) = 0.377 * 0.252 + 0.215 * 0.748 \approx 0.256$$

---

$$P(\neg unemployed) = 1 - P(unemployed) \approx 0.744$$

---

$$P(unemployed|young) = 0.377$$

---

$$P(young|unemployed) = \frac{P(unemployed|young)P(young)}{P(unemployed)} = \frac{0.377 * 0.252}{0.256} \approx 0.371$$

---

$$P(unemployed|\neg young) = 0.215$$

---

$$P(\neg young|unemployed) = \frac{P(unemployed|\neg young)P(\neg young}{P(unemployed)} = \frac{0.215 * 0.748}{0.256} \approx 0.629$$

## 5. Inference

Let's call the event of a child being able to reading at grade 1 as `R` and children who had training as `T`. Then, the probability that a child can read, given that he/she had training (like in case of Olivia) would be:

$$P(R|T) = \frac{P(T|R) * P(R)}{P(T)} = \frac{P(T|R) * P(R)}{P(T|R)P(R) + P(T|\neg R)P(\neg R)} = \frac{0.65 * 0.33}{0.65 * 0.33 + 0.1 * (1 - 0.33)} \approx 76.19\%$$

Using the Bayes' rule, we rewrite the probability in terms of probability of training given reading and then marginalize `P(T)` to make use of the information from educational experts.

---

We have a prior **independent** variable `P(R)` that is a belief on what is the probability. We then update it with data that we get. In this case, we know `P(T|R)` and `P(T|¬R)` as given **independent** variables. The variable we're finding, `P(R|T)`, is a posterior **dependent** variable.