recommended reading on probability:

bishop book, chapter 1 (introduction) and 2 (probability distributions). there is much more in there than just probability theory. very good introduction to the general ideas of machine learning and a lot of stuff (e.g. on gaussians) that we will reuse for later topics

OR

murphy book, chapter 1 (introduction) and 2 (probability). there is much more in there than just probability theory. very good introduction to the general ideas of machine learning. the probability chapter is probably the text most concisely focused on probability theory's relevant aspects for ml.

OR

paper “Review of Probability Theory” by Arian Meleki and Tom Do. The paper that was used as a basis for the slides.

Problem 1: Suppose that a system contains a certain type of component whose time, in years, to failure is given by T. The random variable T is modelled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

Problem 1: Suppose that a system contains a certain type of component whose time, in years, to failure is given by T. The random variable T is modelled nicely by the exponential distribution with mean time to failure $\beta = 5$. If 5 of these components are installed in different systems, what is the probability that at least 2 are still functioning at the end of 8 years?

If the mean of the exponential distribution is $\beta = 5$, its parameter is the inverse $\lambda = \beta^{-1} = 0.2$.

The probability that a given component is still functioning after 8 years is given by

$$p(T > 8) = \int_8^\infty p(t)\,\mathrm{d}t = \frac{1}{5}\int_8^\infty \exp\left(-\frac{x}{5}\right)\,\mathrm{d}t = e^{-8/5} \approx 0.2.$$

Let X represent the number of components functioning after 8 years. Then using the binomial distribution, we have

$$p(X \geq 2) = \sum_{x=2}^{5} \mathrm{Bin}(x \mid 5, 0.2) = 1 - \sum_{x=0}^{1} \mathrm{Bin}(x \mid 5, 0.2) = 1 - 0.7373 = 0.2627.$$

Notice the equivalence of integration and summation for continuous and discrete variables, respectively.

Problem 2: A multiple-choice exam has 200 questions. Each question has four possible answers, with one correct only. What is the probability that random guessing yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge?

You may find it useful to use the *central limit theorem* to solve this exercise:

Let $X_i, i = 1, \ldots, n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \to \mathcal{N}\left(0, \sigma^2\right).$$

(For the mathematically inclined: the convergence is *in distribution.*)

Problem 2: A multiple-choice exam has 200 questions. Each question has four possible answers, with one correct only. What is the probability that random guessing yields from 25 to 30 correct answers for the 80 of the 200 problems about which the student has no knowledge?

You may find it useful to use the *central limit theorem* to solve this exercise:

Let $X_i, i = 1, \ldots, n$ be a sequence of i.i.d. random variables with mean μ and variance σ^2. Then

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \to \mathcal{N}\left(0, \sigma^2\right).$$

(For the mathematically inclined: the convergence is *in distribution*.)

The probability of guessing a correct answer for each of the 80 questions is $\theta = 1/4$. If X represents the number of correct answers resulting from guesswork, then

$$p(25 \leq X \leq 30) = \sum_{x=25}^{30} \mathrm{Bin}(x \mid 80, 1/4).$$

At this point, we would only need to sum up these six values. The problem is that it involves factorials of very large numbers. These are hard to compute with limited precision, in particular since we subsequently need fractions of them. Luckily, a binomial variable is a sum of Bernoulli variables, $X = \sum_{i=1}^{n} X_i$, and $n = 80$ is high enough for the central limit theorem to kick in. Using $\mu = \theta$ and $\sigma^2 = \theta(1 - \theta)$, we get

$$
\begin{aligned}
&p(25 \leq X \leq 30) \\
=&p\left(\frac{25}{n} \leq \frac{X}{n} \leq \frac{30}{n}\right) \\
=&p\left(\frac{25-n\mu}{n} \leq \frac{X-n\mu}{n} \leq \frac{30-n\mu}{n}\right) \\
=&p\left(\frac{25-n\mu}{\sqrt{n}} \leq \frac{X-n\mu}{\sqrt{n}} \leq \frac{30-n\mu}{\sqrt{n}}\right).
\end{aligned}
$$

We observe that the random variable in the middle now corresponds to the one in the central limit theorem, which converges to a Gaussian distribution with variance $\theta(1-\theta)$. Hence, if we divide by the standard deviation $\sqrt{\theta(1-\theta)}$, we get a *standard* Gaussian distribution:

$$
p(25 \leq X \leq 30) = p\left(\frac{25-n\theta}{\sqrt{n\theta(1-\theta)}} \leq \underbrace{\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}}}_{\approx Z \sim \mathcal{N}(0,1)} \leq \frac{30-n\theta}{\sqrt{n\theta(1-\theta)}}\right)
$$

One more thing: The binomial distribution does not have any mass between, e.g., 24 and 25. At the same time, the Gaussian distribution does have mass between 24 and 25. To account for that, we change 25 to 24.5 and 30 to 30.5 (by convention).

With $n = 80$ and $\theta = 0.25$ and Φ the cdf of the standard Gaussian distribution:

$$
\begin{aligned}
p(25 \leq X \leq 30) &\approx \Phi\left(\frac{30.5-n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \Phi\left(\frac{24.5-n\theta}{\sqrt{n\theta(1-\theta)}}\right) \\
&= \Phi(2.71) - \Phi(1.16) = 0.1196
\end{aligned}
$$

The area of interest can be seen as a blue shade in Fig. 1.

If you do the exact calculations from the original binomial distribution (taking care of numerical precision), you get approximately 0.11927—pretty close! This shows one of the intriguing properties of normal distributions.

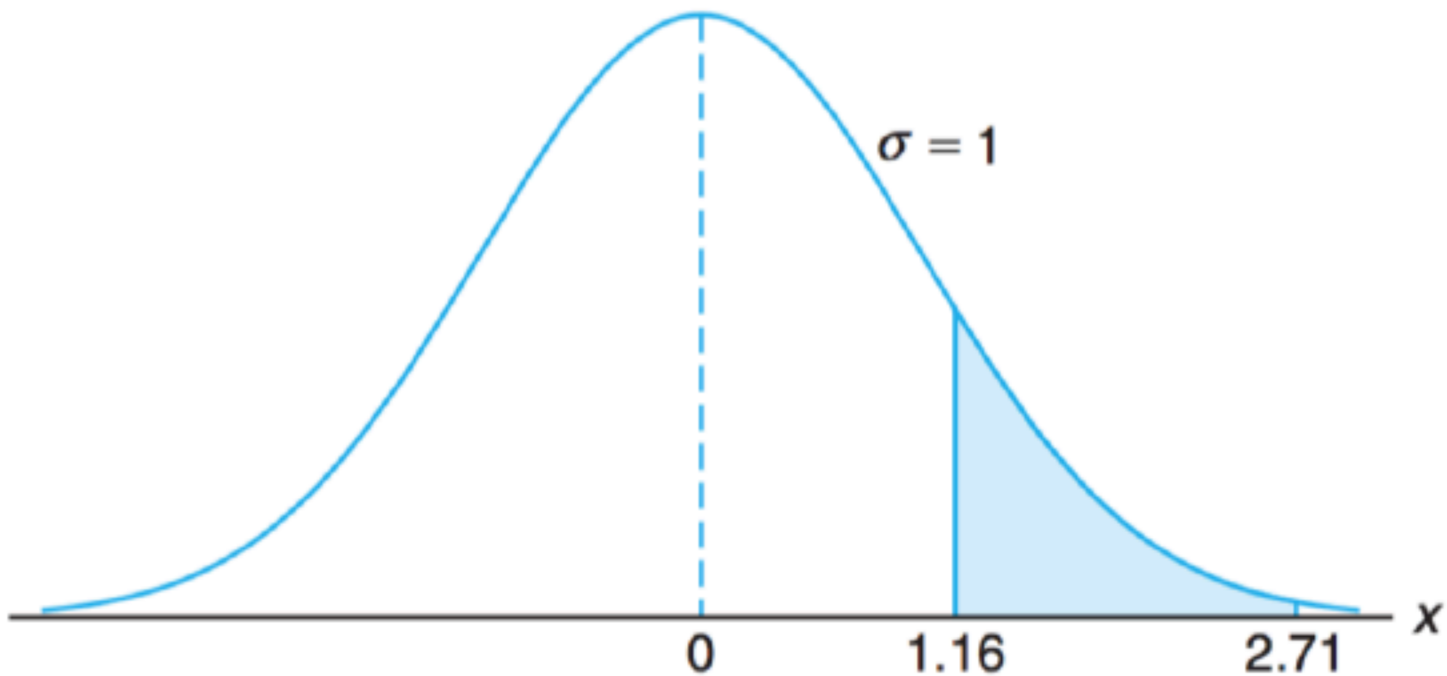

Figure 1: For first tutorial question.

Problem 3: As an avid player of board games you have a nice collection of non-standard dice: You have a 3-sided, 5-sided, 7-sided, 11-sided and 20-sided die. The five dice are treasured in a beautiful purple velvet bag. Without looking, a friend of yours randomly chooses a die from the bag and rolls a 6. What is the probability that the 11-sided die was chosen? What is the probability that the 20-sided die was used for the role? Show your work!

Now your friend rolls (with the same die!) an 18. What is the probability now that the die is 11-sided? What is the probability that it is 20-sided? Show your work!

Problem 3: As an avid player of board games you have a nice collection of non-standard dice: You have a 3-sided, 5-sided, 7-sided, 11-sided and 20-sided die. The five dice are treasured in a beautiful purple velvet bag. Without looking, a friend of yours randomly chooses a die from the bag and rolls a 6. What is the probability that the 11-sided die was chosen? What is the probability that the 20-sided die was used for the role? Show your work!

Now your friend rolls (with the same die!) an 18. What is the probability now that the die is 11-sided? What is the probability that it is 20-sided? Show your work!

Let the random variable N denote the number of sides of the die, and let the random variable R denote the result of the roll.

We are interested in the probability $p(N = 11 \mid R = 6)$. By Bayes' theorem, we know

$$p(N = 11 \mid R = 6) = \frac{p(R = 6 \mid N = 11)p(N = 11)}{p(R = 6)}.$$

We know that $p(N = n) = 0.2$, as we have a random fair draw from 5 dice. Furthermore

$$p(R = r \mid N = n) = \begin{cases} \frac{1}{n} & 1 \le r \le n \\ 0 & \text{else.} \end{cases}$$

This determines the numerator. For the denominator, we use the law of total probability:

$$p(R = r) = \sum_{n \in \{3,5,7,11,20\}} p(R = r \mid N = n) p(N = n)$$

Inserting $r = 6$ and $n = 11$ or $n = 20$ yields posterior probabilities of 0.32036 and 0.17620, respectively. Notice the eleven-sided die is more likely because it is more likely to produce a 6, even though die draw and roll are uniformly random experiments!

We can also insert $r = 18$, which gives us a probability of 0 for $n = 11$ and 1 for $n = 20$, which matches our intuition, as the 20-sided die is the only one which can produce an 18.

Problem 4: Suppose that 15 percent of the items produced in a certain plant are defective. If an item is defective, the probability is 0.9 that it will be moved to the waste. If an item is not defective, the probability is 0.2 that it is moved (accidentally) to the waste.

If an item is removed, what is the probability that it is defective?

Problem 4: Suppose that 15 percent of the items produced in a certain plant are defective. If an item is defective, the probability is 0.9 that it will be moved to the waste. If an item is not defective, the probability is 0.2 that it is moved (accidentally) to the waste.

If an item is removed, what is the probability that it is defective?

Let D denote the part being defective, let M denote the part being moved to wasted, then

$$p(D \mid M) = \frac{p(M \mid D)p(D)}{p(M)} = \frac{0.9 \cdot 0.15}{0.9 \cdot 0.15 + 0.2 \cdot 0.85} \approx 0.44.$$

Problem 5: A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$p(D \mid P_1) = 0.01, \quad p(D \mid P_2) = 0.03, \quad p(D \mid P_3) = 0.02,$$

where $p(D \mid P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

Problem 5: A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$p(D \mid P_1) = 0.01, \quad p(D \mid P_2) = 0.03, \quad p(D \mid P_3) = 0.02,$$

where $p(D \mid P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

From the statement of the problem

$$p(P_1) = 0.30, \quad p(P_2) = 0.20, \quad \text{and} \quad p(P_3) = 0.50,$$

we must find $p(P_j \mid D)$ for $j = 1, 2, 3$. Bayes' rule shows

$$\begin{aligned} p(P_1 \mid D) &= \frac{p(P_1)p(D \mid P_1)}{p(P_1)p(D \mid P_1) + p(P_2)p(D \mid P_2) + p(P_3)p(D \mid P_3)} \\ &= \frac{0.30 \cdot 0.01}{0.3 \cdot 0.01 + 0.20 \cdot 0.03 + 0.50 \cdot 0.02} = \frac{0.003}{0.019} = 0.158. \end{aligned}$$

Similarly,

$$p(P_1 \mid D) = \frac{0.03 \cdot 0.20}{0.019} = 0.316 \quad \text{and} \quad p(P_3 \mid D) = \frac{0.02 \cdot 0.50}{0.019} = 0.526.$$

The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3.

Problem 6: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x, y) = \begin{cases} 8xy & 0 \le y \le x \le 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Problem 6: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x, y) = \begin{cases} 8xy & 0 \leq y \leq x \leq 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

From the lecture, we know that $\text{Cov}(X, Y) = E[X, Y] - E[X]E[Y]$. Hence, to compute $E[X]$ and $E[Y]$, we first compute the marginal density functions. The marginal distribution of X is

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}y = \int_0^x 8xy\, \mathrm{d}y = \begin{cases} 4x^3, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

We dropped the region outside the interval $[0, x]$ in the integral as the density is zero there.

Similarly,

$$h(y) = \int_y^1 8xy\, \mathrm{d}x = \begin{cases} 4y(1 - y^2), & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Again, we dropped the region where the density is 0 for a given y.

From these marginal density functions, we compute

$$E(X) = \int_0^1 xg(x)\,\mathrm{d}x = \int_0^1 4x^4\,\mathrm{d}x = \frac{4}{5}$$
$$E(Y) = \int_0^1 yg(y)\,\mathrm{d}y = \int_0^1 4y^2(1-y^2)\,\mathrm{d}y = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2\,\mathrm{d}x\,\mathrm{d}y = \frac{4}{9}.$$

Then

$$\mathrm{Cov}(X,Y) = E[X,Y] - E[X]E[Y] = \frac{4}{9} - \frac{4}{5}\frac{8}{15} = \frac{4}{225}.$$