CHAPTER 1

EXERCISES: BASIC CONCEPTS IN PROBABILITY AND STATISTICS

Table 1.1 Sample Outcome of 100 pairs of coin tosses

	H_2	T_2
H_1	30	21
T_1	22	27

Exercise 1.1 (Empirical Probabilities)**.** The *relative frequency* of an event is the fraction of times that the event occurs. Suppose a coin is tossed 200 times. The outcome of 100 consecutive pairs of tosses is summarized in the contingency table 1.1, where H_1 and T_1 indicate the number of heads and tails on the first toss, respectively, and H_2 and T_2 indicate the number of heads and tails on the second toss. For instance, the "21" listed in the top right entry of the table means that 21 out of 100 pairs was "heads-tails." The empirical probability of "heads-tails" is therefore 21/100.

a) What is the empirical probability of heads on the first toss of a pair, denoted $P(H_1)$?

b) What is the empirical probability of a heads in the second toss of a pair, denoted $P(H_2)$?

c) What is the empirical probability of the joint event H_1 and T_2, denoted $P(H_1, T_2)$?

d) What is the empirical probability of T_2 *given* H_1, denoted $P(T_2|H_1)$? Compute this by examining only those cases in which H_1 is true.

e) Verify that the empirical conditional probability computed in (d) is the ratio of the joint probability over the marginal: $P(T_2|H_1) = P(T_2, H_1)/P(H_1)$.

(Incidentally, this exercise suggests why unconditional probabilities are called marginal probabilities, namely because they can be calculated by summing values in a table along rows and columns and writing the results along the margins of the table.) □

Exercise 1.2. If X and Y are independent, then show that $p(x, y) = p(x)p(y)$ implies $E[XY] = E[X]E[Y]$. □

Exercise 1.3. Let X and Y be independent random variables with respective means μ_X and μ_Y, and respective variances σ_X^2 and σ_Y^2. In terms of the population means and variances:

(a) Compute $E[XY]$?

(b) Compute $\text{cov}[X, Y]$?

(c) Compute $\text{var}[X - Y]$?

(d) Compute $\text{var}[XY]$?

□

Exercise 1.4 (Properties of Variance)**.** Let k be a constant and X be a random variable with expectation μ_X and variance σ_X^2. Using the definition of variance and the properties of expectation, prove the following:

(a) $\text{var}[k] = 0$

(b) $\text{var}[kX] = k^2 \,\text{var}[X]$?

(c) $\text{var}[k + X] = \text{var}[X]$?

□

Exercise 1.5. Using the properties of expectation, show that

$$\text{cov}[X, Y] = E[XY] - E[X]E[Y]. \tag{1.1}$$

□

Exercise 1.6. If X is a random variable and k is a constant (i.e., independent of X), then show that $E[(X - k)^2]$ is minimized when $k = E[X]$. (Hint: there are at least two ways to prove this: one way is to use calculus, the other way is to add and subtract an expectation. Can you find both proofs?) Be sure to prove that the solution is actually a *minimum*. □

Exercise 1.7 (Bounds on the Correlation Coefficient)**.** Prove that the correlation coefficient between any two random variables X and Y is always between -1 and 1. Explain when the correlation is exactly equal to one, and when the correlation is exactly equal to -1, in terms of the relation between X and Y. Hint: use the fact that

$$E[(t(X - E[X]) + (Y - E[Y]))^2] = t^2 \,\text{var}[X] + 2t \,\text{cov}[X, Y] + \text{var}[Y] \geq 0, \tag{1.2}$$

for any t. Note that this equation for t is a parabola. What is the minimum value? □

Exercise 1.8 (Distribution of a Difference in Means)**.** Suppose $X_1, X_2, \ldots, X_N$ are independent random random variables drawn from a Gaussian distribution with mean μ_X and variance σ^2. Similarly, let $Y_1, Y_2, \ldots, Y_N$ be identically distributed random variables drawn from a Gaussian distribution with different mean μ_Y but common variance σ^2. Also, assume that the X's and Y's are independent.

- What are the distributions of the sample means $\hat{\mu}_X$ and $\hat{\mu}_Y$? (Do not just say "normal." Tell me all the parameters– e.g., the mean and variance.)
- What is the distribution of the difference in sample means $\hat{\mu}_X - \hat{\mu}_Y$?
- What is the distribution of the sample variances $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$? (Tell me all parameters.)
- What is the distribution of the sum of sample variances $\hat{\sigma}_X^2 + \hat{\sigma}_Y^2$?

□

Exercise 1.9 (Expectation of a Roll of a Die)**.** If X is the outcome of the roll of a fair die, what is $E[X]$? What is $\mathrm{var}[X]$? □

Exercise 1.10 (Expectations on Dice)**.** Let X be the *sum* of the outcome of the roll of three fair dice. What is $E[X]$? What is $\mathrm{var}[X]$? Assume the outcome of any individual die is independent of that of the others. □

Exercise 1.11 (Prove the Sample Variance is Unbiased)**.** Let $X_1, \ldots, X_N$ be independent random samples drawn from a population with expectation μ and variance σ^2. Show that the expectation of the sample variance $\hat{\sigma}_X^2$ is

$$E[\hat{\sigma}_X^2] = \sigma^2 \tag{1.3}$$

For this reason, $\hat{\sigma}_X^2$ is called an *unbiased* estimate of σ_X^2. (This question is hard) □

Exercise 1.12. Suppose you are predicting whether Y will be above or below normal. If "normal" is defined as the median, then, by definition, Y will be above normal 50% of the time (like a coin flip). Suppose, however, you are predicting Y given knowledge of another variable X. If X and Y are independent, then knowledge of X tells you nothing new about Y; i.e., the probability that Y is above normal is still 50%. However, if X and Y have a non-zero correlation, then knowledge of one tells you something new about the other. Suppose that the correlation between X and Y is 0.5. An example of such a sample is shown in fig. 1.3c. If X is above normal, what is the probability that Y is above normal? To answer this question, use example 1.7 to generate random numbers with a population correlation of $\rho = 0.5$, and then from these numbers compute the probability that $Y > 0$ given that $X > 0$. You should generate enough random samples to obtain *robust* estimates of the probability (i.e., ensure your answer does not change if you re-run your code with different random numbers).

□

Exercise 1.13 (Random Walk in Two Dimensions)**.** A particle takes random steps in the xy-plane. The step increment in the x-direction is determined by drawing independent random numbers from a normal distribution with zero mean and variance σ^2. Positive values correspond to steps in the positive x-direction and negative values correspond to steps in the negative x-direction. The step increment in the y-direction is determined in the analogous way. Find the equation for the radius of the circle such that there is a 95% probability that the particle lies within the circle after 10 steps. Evaluate this equation for the case $\sigma = 2$. □

Exercise 1.14. In exercise 1.6, you proved that $k = \mathbb{E}[X]$ minimized $\mathbb{E}[(X-k)^2]$. This says that the best prediction of X (in a mean square sense) is the mean. However, the population mean is unknown. Accordingly, consider a prediction of X based on the sample mean $\hat{\mu}_X$, derived from $X_1, \dots, X_N$. It turns out that the sample mean is *not* the best prediction of X', when X' is independent of the sample used to construct $\hat{\mu}_X$. Introduce a scaling factor α, and show that the α that minimizes

$$\mathbb{E}[(X' - \alpha\hat{\mu})^2],$$

is

$$\alpha = \frac{1}{\left(\frac{\sigma_X^2}{N\mu_X^2}\right) + 1},$$

where μ_X and σ_X^2 are the population means and variances of X. Show that $\alpha < 1$, which explains why this predictor is sometimes called a *shrinkage* estimator. □