46-880 Introduction to Probability and Statistics, mini-1 2023
Solution to Little Test 1

1. Suppose you roll three dice. What is the probability that at least two of the rolls are the same?

   Solution. It is easier to compute the complement:

   $$\mathbb{P}(\text{the three rolls are different}) = \frac{6 \cdot 5 \cdot 4}{6^3} = \frac{20}{36}.$$

   Hence

   $$\mathbb{P}(\text{at least two rolls are the same}) = 1 - \mathbb{P}(\text{the three rolls are different}) = 1 - \frac{20}{36} = \frac{16}{36}.$$

2. Suppose 80% of all statisticians are shy, whereas only 15% of economists are shy. Suppose that 90% of the people at a large gathering are economists and the other 10% are statisticians.

   You meet a random person at the gathering. If that person is shy, what is the probability that the person is a statistician?

   Solution.

   Consider the following events

   A : the random person is a statistician.

   B : the random person is shy.

   We know $\mathbb{P}(A) = 0.1$ and also $\mathbb{P}(B|A) = 0.8$ and $\mathbb{P}(B|A^c) = 0.15$. We want $\mathbb{P}(A|B)$. Apply Bayes' Theorem:

   $$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)} = \frac{0.8 \cdot 0.1}{0.8 \cdot 0.1 + 0.15 \cdot 0.9} = \frac{0.08}{0.215} = 0.372$$

3. Suppose 80% of all statisticians are shy, whereas only 15% of economists are shy. Suppose that 90% of the people at a large gathering are economists and the other 10% are statisticians.

   You meet a random person at the gathering. What is the probability that the person is shy?

   Solution. We want

   $$\mathbb{P}(B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c) = 0.8 \cdot 0.1 + 0.15 \cdot 0.9 = 0.215.$$

4. Suppose X is a binomial random variable with 50 trials and probability of success 0.44, that is, $X \sim B(50, 0.44)$.

   What is the numerical value that X takes with highest probability?

   Solution. The possible values of X are $0, 1, \cdots, 50$. Using a spreadsheet we can compute

   $$\mathbb{P}(X = x) = \texttt{binom.dist}(\texttt{x}, 50, 0.44, 0) \text{ for } x = 0, 1, \ldots, 20.$$

   It is then evident that the highest probability value is 22 with probability

   $$\texttt{binom.dist}(22, 50, 0.44, 0) = 0.113.$$

5. A hotel has 50 rooms. Assume that a reservation is a "no-show" (that is, it does not show up) with probability 0.1. Consequently, the hotel routinely accepts more than 50 reservations. What is the largest number of reservations that the hotel can accept so that the probability that the hotel will be overbooked is less than 0.2?

   (The hotel is overbooked if it does not have enough rooms for the reservations who show up.)

   Solution. Let n = number of reservations that the hotel accept and X = number of reservations that show up. Then $X \sim B(n, 0.9)$ and the hotel is overbooked when $X \geq 51$. Hence the probability that the hotel is overbooked is

$$\mathbb{P}(X \geq 51) = 1 - \mathbb{P}(X \leq 50) = 1 - \texttt{binom.dist}(50, \texttt{n}, 0.9, 1).$$

   We have the following values of this probability for different values of n

n	51	52	53	54	55	56
$\mathbb{P}(X \geq 51)$	0.00464	0.02829	0.0898	0.1985	0.3451	0.5065

   Thus the largest n can be so that $\mathbb{P}(X \geq 51) < 0.2$ is $n = 54$.

6. Suppose calls arrive at a 1-800 service line according to a Poisson process with an average rate of arrival of 12 calls per hour.

   What is the probability that at least one call arrives within the next 10 minutes?

   Solution. Let X = number of calls that arrive in the next 10 minutes. Thus $X \sim \text{Pois}(12/6)$ and we want $\mathbb{P}(X \geq 1)$. That is

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \texttt{poisson.dist}(0, 12/6, 0) = 0.86466.$$

Solution to Little Test 2

1. Suppose X is a random variable with cumulative distribution function $F(x)$. If a and b are numbers with $a \leq b$, then $F(a) \leq F(b)$.

   Solution. TRUE: Since $\{X \leq a\} \subseteq \{X \leq b\}$ it follows that

   $$F(a) = \mathbb{P}(X \leq a) \leq \mathbb{P}(X \leq b) = F(b).$$

2. Suppose X is a continuous random variable with density function $f(x)$. If a and b are numbers with $a \leq b$, then $f(a) \leq f(b)$.

   Solution. FALSE: Suppose $X \sim U(0,1)$, that is, X is uniformly distributed between 0 and 1. Then $f(0.5) = 1$ but $f(2) = 0 < f(0.5)$.

3. If the random variables X, Y are uncorrelated then $\mathrm{var}(X+Y) = \mathrm{var}(X-Y)$ then X, Y.

   Solution. TRUE: We know that

   $$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$

   and

   $$\mathrm{var}(X-Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\mathrm{cov}(X,Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

   Hence $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) = \mathrm{var}(X-Y)$.

4. If the random variables X, Y are such that $\mathrm{var}(X+Y) = \mathrm{var}(X-Y)$ then X, Y are uncorrelated.

   Solution. TRUE: We know that

   $$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$$

   and

   $$\mathrm{var}(X-Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\mathrm{cov}(X,Y).$$

   Hence $\mathrm{var}(X+Y) = \mathrm{var}(X-Y)$ implies that

   $$\mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y) = \mathrm{var}(X) + \mathrm{var}(Y) - 2\mathrm{cov}(X,Y).$$

   Hence $4\mathrm{cov}(X,Y) = 0$ and thus $\mathrm{cov}(X,Y) = 0$. Therefore X, Y are uncorrelated.

5. Suppose X is a normal random variable such that 0.995-quantile(X) $= 40.909$ and 0.975-quantile(X) $= 33.519$.

   Find the mean of X.

   Solution. Suppose $\mu = \mathbb{E}(X)$ and $\sigma = \mathrm{stdev}(X)$. From the properties of the normal distribution we know that

   $$0.995\text{-quantile}(X) = \mu + 2.576\sigma$$

and

$$0.975\text{-quantile}(X) = \mu + 1.96\sigma.$$

Hence we have the two equations

$$\begin{aligned} \mu + 2.576\sigma &= 40.909 \\ \mu + 1.96\sigma &= 33.519 \end{aligned}$$

If we multiply the first one by 1.96 and the second one by 2.576 and subtract, we get

$$0.616\mu = (2.576 - 1.96)\mu = 2.576 \cdot 33.519 - 1.96 \cdot 40.909 = 6.1633.$$

Thus $\mu = 10.0536$.

6. Suppose the random variables $X_1, \ldots, X_n$ are iid standard normal and consider their sample mean

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Find the smallest integer n such that $|\bar{X}_n| \leq 0.1$ with probability at least 0.95. In other words, find the smallest integer n such that $\mathbb{P}(|\bar{X}_n| \leq 0.1) \geq 0.95$.

Solution. The properties of normal distributions, iid, and sample mean imply that $\bar{X}_n \sim N(0, 1/n)$ or equivalently $\sqrt{n}\bar{X}_n \sim N(0,1)$. Thus we have

$$\mathbb{P}(|\bar{X}_n| \leq 0.1) = \mathbb{P}(|\sqrt{n}\bar{X}_n| \leq 0.1\sqrt{n}).$$

We know that for $Z \sim N(0,1)$ we have $\mathbb{P}(|Z| \leq 1.96) = 0.95$. Since $\sqrt{n}\bar{X}_n \sim N(0,1)$ to have $\mathbb{P}(|\sqrt{n}\bar{X}_n| \leq 0.1\sqrt{n}) \geq 0.95$ we must have $0.1\sqrt{n} \geq 1.96$, that is

$$n \geq 19.6^2 = 384.16.$$

Thus $n = 385$.

Solution to Little Test 3

1. Suppose a population is normally distributed with mean 20 and variance 4. Compute a 90% acceptance interval for the sample mean of random samples of size 36.

   Solution. The $(1-\alpha)$ acceptance interval is

   $$\mu \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

   We have $\mu = 20, \sigma = 2, \alpha = 0.1$, and $n = 36$. Thus the 90% acceptance interval is

   $$20 \pm 1.6448 \cdot \frac{2}{\sqrt{36}} = 20 \pm 0.54828$$

2. Suppose a poll of 3000 American adults reveals that 1400 of them approve of the job that Joe Biden is doing as president. Thus the estimate of the proportion of the population who approve Joe Biden is $1400/3000 = 0.467$

   Find the margin of error of this estimate at the 0.95 confidence level. (That is, the radius of the 0.95 confidence interval for the population proportion.)

   Please enter the number with four digits of accuracy. Note that the number is between 0.0001 and 0.1000.

   Solution. We have $n = 3000$ and $\hat{p} = 1400/3000 = 0.467$. Thus the MOE at the 0.95 confidence level is

   $$z_{0.025} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{3000}} = 1.96 \cdot \sqrt{\frac{0.467 \cdot 0.533}{3000}} = 0.017853$$

3. A hypothesis test for population mean based on a random sample of size n rejects the null hypothesis at the significance level if the p-value of the test is less than α.

   Solution. TRUE: This is one of the rules of the t-test.

4. Suppose the sample correlation of 200 observations of the variables X and Y is 0.2. Then we reject the null hypothesis that the correlation of X and Y is zero at the 0.05 significance level.

   Solution. TRUE: Compute the t-stat. We have $n = 200$ and $r = 0.2$. Hence

   $$\text{t-stat} = 0.2 \cdot \sqrt{\frac{198}{1 - 0.2^2}} = 2.8722$$

   which is larger the the critical t-value:

   $$t_{198,0.025} = \texttt{T.INV}(0.975, 198) = 1.972.$$

5. A hypothesis test for population mean based on a random sample of size n rejects the null hypothesis at the significance level α if |t-stat| is less than $t_{n-1,\alpha/2}$.

   Solution. FALSE: The rule for the t-test is that we reject at the significance level α if |t-stat| is LARGER than $t_{n-1,\alpha/2}$.

6. The estimate of a regression coefficient is statistically significant at level α when the $(1-\alpha)$ confidence interval of that coefficient does not contain the value zero.

   Solution. TRUE: The estimate is statistically significant at level α when we reject the null hypothesis that it is zero at level α. The above statement is one of the rejection rule for the t-test.