Matthew Wilder
MATH 326 - Spring 2022
Homework 01
Due: Wednesday, 01/26/22 23:59

1. Benford's Law states that in a legitimate financial record, 30.1% of all randomly selected first digits will be "one".

   (a) What is the probability that exactly two out of 10 records begin with a "1"?

   Solution: This is a binomial distribution with

   $$p = 0.301 \qquad q = 0.699 \qquad n = 10 \qquad y = 2$$

   Substituting into the binomial distribution formula, we obtain the answer

   $$\begin{aligned} P(Y=2) &= f(2) \\ &= \binom{10}{2} p^2 q^{10-2} \\ &= \binom{10}{2} (0.301)^2 (0.699)^8 \\ &\approx 45 \cdot 0.090601 \cdot 0.05699 \\ &\boxed{\approx 0.23236} \end{aligned}$$

   (b) What is the probability that at least 20 out of 100 records begin with a "1"?

   Solution: This is $P(Y \geq 20)$ which is equivalent to $1 - P(Y < 20)$ Similar to part (a),

   $$p = 0.301 \qquad q = 0.699 \qquad n = 100 \qquad y = k \qquad \text{for} \quad k \in [0, 100] \cap \mathbb{Z}$$

   $$\begin{aligned} P(Y \geq 20) &= 1 - P(Y < 20) \\ &= 1 - \sum_{k=0}^{k=19} f(k) \\ &= 1 - \sum_{k=0}^{k=19} \left[\binom{100}{k} (0.301)^k (0.699)^{100-k} \right] \\ &\boxed{\approx 0.991605} \qquad \text{by Wolfram Alpha} \end{aligned}$$

   (c) What is the probability that we have not seen a "1" in our first eight records?

   Solution: This is, again, a binomial distribution with

   $$p = 0.301 \qquad q = 0.699 \qquad n = 8 \qquad y = 0$$

Where we want to find $P(Y = 0)$. Applying the binomial distribution formula once again, we obtain

$$\begin{aligned} P(Y = 0) &= f(0) \\ &= \binom{8}{0} p^0 q^{8-0} \\ &= 1 \cdot 1 \cdot (0.699)^8 \\ &\boxed{\approx 0.05699246} \end{aligned}$$

2. Consider a continuous random variable Y with density function

$$f(y) = \frac{k}{y} \text{ with support } 1 \leq y \leq 3.$$

(a) Find the value of k that will make $f(y)$ a legitimate density function.

Solution: In order for f to be a valid probability density function, $\int_S f = 1$ on support $S \in [1, 3]$. Therefore, we need $\int_1^3 f = 1$. We will integrate with the fixed constant k and solve for it after evaluating.

$$\begin{aligned}
\int_1^3 f &= \int_1^3 \frac{k}{y}\,dy \\
&= k \int_1^3 \frac{1}{y}\,dy \\
&= k \Big[\ln(y) \Big]_{y=1}^{y=3} \\
&= k \Big[\ln(3) - \ln(1) \Big] \\
&= k \ln 3 \\
&= 1
\end{aligned}$$

Now, solving the equation $k \ln 3 = 1$ for k, we see that

$$\boxed{k = \frac{1}{\ln 3} \approx 0.910239}$$

(b) Find $P(Y \leq 2.5)$.

Solution: Filling in the information obtained from part (a), we see that

$$f(y) = \begin{cases} \frac{1}{y \ln 3} & y \in [1, 3] \\ 0 & \text{elsewhere} \end{cases}$$

$$
\begin{aligned}
P(Y \leq 2.5) &= P(-\infty < Y \leq 2.5) \\
&= \int_{-\infty}^{2.5} f \\
&= \int_{-\infty}^{1} f + \int_{1}^{2.5} f \\
&= 0 + \int_{1}^{2.5} \frac{1}{y \ln 3} \\
&= \frac{1}{\ln 3} \int_{1}^{2.5} \frac{1}{y} \\
&= \frac{1}{\ln 3} \Big[\ln y \Big]_{y=1}^{y=2.5} \\
&= \frac{\ln 2.5 - \ln 1}{\ln 3} \\
&= \frac{\ln 2.5}{\ln 3} \\
&\boxed{\approx 0.834043767}
\end{aligned}
$$

(c) Find the mean and standard deviation of Y.

Solution: Using the definition of expected value on a continuous pdf, $E[Y] = \int_S yf$, we can compute the following

$$
\begin{aligned}
E[Y] &= \int_S yf \\
&= \int_1^3 \frac{y}{y \ln 3} dy \\
&= \frac{1}{\ln 3} \int_1^3 dy \\
&= \frac{3-1}{\ln 3} \\
&= \frac{2}{\ln 3} \\
&\approx 1.82047845
\end{aligned}
$$

Similarly we can compute variance from the definition $V[Y] = E[Y^2] - \big(E[Y]\big)^2$

$$\begin{aligned}
V[Y] &= E[Y^2] - \big(E[Y]\big)^2 \\
&= \int_S y^2 f - \left(\frac{2}{\ln 3}\right)^2 \\
&= \int_1^3 \frac{y^2}{y\ln 3}dy - \frac{4}{\ln^2 3} \\
&= \int_1^3 \frac{y}{\ln 3}dy - \frac{4}{\ln^2 3} \\
&= \frac{1}{\ln 3}\int_1^3 ydy - \frac{4}{\ln^2 3} \\
&= \frac{1}{2\ln 3}\Big[y^2\Big]_{y=1}^{y=3} - \frac{4}{\ln^2 3} \\
&= \frac{9-1}{2\ln 3} - \frac{4}{\ln^2 3} \\
&= \frac{4}{\ln 3} - \frac{4}{\ln^2 3} \\
&= 4\left(\frac{1}{\ln 3} - \frac{1}{\ln^2 3}\right) \\
&= 4\left(\frac{\ln 3 - 1}{\ln^2 3}\right) \\
&\approx 0.326815108
\end{aligned}$$

Since $V[Y] = \sigma^2$, the standard deviation, σ is

$$\sigma = \sqrt{\sigma^2} = \sqrt{V[Y]} = \sqrt{4\left(\frac{\ln 3 - 1}{\ln^2 3}\right)} = \frac{2\sqrt{\ln 3 - 1}}{\ln 3} \approx 0.571677451$$

$$\boxed{\underbrace{\mu = \frac{2}{\ln 3} \approx 1.82047845}_{\text{mean}} \qquad \underbrace{\sigma = \frac{2\sqrt{\ln 3 - 1}}{\ln 3} \approx 0.571677451}_{\text{standard deviation}}}$$

(d) Suppose random variables Y_1, Y_2, Y_3, and Y_4 are independent random event from the above distribution $f(y)$. Let $M = \max(Y_1, Y_2, Y_3, Y_4)$. Find $P(M \leq 2.5)$.

Solution: We want to find $P\big(\max(Y_1, Y_2, Y_3, Y_4) \leq 2.5\big)$, which is equivalent to $P\Big((Y_1 \leq 2.5) \wedge (Y_2 \leq 2.5) \wedge (Y_3 \leq 2.5) \wedge (Y_4 \leq 2.5)\Big)$. Since Y is an i.i.d., this is further equivalent to $P\left((Y \leq 2.5)^4\right)$.

That is to say, what is the probability that *all* 4 individual (identical) events are successes (less than 2.5). This can be written as a binomial distribution with

$$p = \frac{\ln 2.5}{\ln 3} \qquad n = 4 \qquad y = 4$$

Substituting into the binomial distribution formula, we get

$$\begin{aligned} P\left(Y=4\right) &= \binom{4}{4} p^4 q^{4-4} \\ &= 1 \cdot p^4 q^0 \\ &= \left(\frac{\ln 2.5}{\ln 3}\right)^4 \\ &\boxed{\approx 0.483899713} \end{aligned}$$

3. Consider the random variable X and Y whose joint probability distribution $p(x, y)$ is given in the following table.

$X \backslash Y$	0	1	2
1	0.15	0.10	0.05
2	0.05	0.20	0.10
3	0.05	0.05	0.25

Find each of the following:

(a) $p_x(1) = P(X = 1)$

Solution: This is going to be the sum of all the probabilities such that $x = 1$. We can fix x at 1 and iterate over the y's. Thus,

$$\begin{aligned} P(X = 1) &= \sum_{y=0}^{2} p(1, y) \\ &= p(1,0) + p(1,1) + p(1,2) \\ &= 0.15 + 0.10 + 0.05 \\ &\boxed{= 0.30} \end{aligned}$$

(b) $E[X]$

Solution: Using the definition of discrete expected value,

$$\begin{aligned} E[X] &= \sum_{x=1}^{3} xP\left(X = x\right) \\ &= 1\left(P\left(X = 1\right)\right) + 2\left(P\left(X = 2\right)\right) + 3\left(P\left(X = 3\right)\right) \\ &= 1\sum_{y=0}^{2} p(1, y) + 2\sum_{y=0}^{2} p(2, y) + \sum_{y=0}^{2} p(3, y) \\ &= 1(0.15 + 0.10 + 0.05) + 2(0.05 + 0.20 + 0.10) + 3(0.05 + 0.05 + 0.25) \\ &= 1(0.30) + 2(0.35) + 3(0.35) \\ &\boxed{= 2.05} \end{aligned}$$

(c) $P(X = 1 \mid Y = 2)$

Solution: We begin with computing the probability that $Y = 2$.

$$p_y(2) = \sum_{x=1}^{3} p(x, 2) = 0.05 + 0.10 + 0.25 = 0.40$$

We have already show in (a) that $p_x(1) = 0.30$. Using the following formula,

$$P(X = x \mid Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

Substituting in,

$$\begin{aligned} P(X = 1 \mid Y = 2) &= \frac{P(X = 1 \cap Y = 2)}{P(Y = 2)} \\ &= \frac{p(1,2)}{p_y(2)} \\ &= \frac{0.05}{0.40} \\ &\boxed{= 0.125} \end{aligned}$$

(d) $E(X \mid Y = 2)$

Solution: Using a modified definition for discrete expected value, where y is fixed,

$$\begin{aligned} E(X \mid Y = 2) &= \sum_{x=1}^{3} xp(x,2) \\ &= 1\,(p(1,2)) + 2\,(p(2,2)) + 3\,(p(3,2)) \\ &= 1(0.05) + 2(0.10) + 3(0.25) \\ &\boxed{= 1} \end{aligned}$$

(e) Find the mean and variance of each random variable.

Solution: We already know $E[X] = 2.05$ from part (b), so our next step would be computing $V[Y]$. Similarly to how we computed it in part (b),

$$\begin{aligned} E[Y] &= \sum_{y=0}^{2} yP(Y = y) \\ &= 0\,(P(Y = 0)) + 1\,(P(Y = 1)) + 2\,(P(Y = 2)) \\ &= 0\sum_{x=1}^{3} p(x,0) + 1\sum_{x=1}^{3} p(x,1) + 2\sum_{x=1}^{3} p(x,2) \\ &= 0 + (0.10 + 0.20 + 0.05) + 2(0.05 + 0.10 + 0.25) \\ &= 1.15 \end{aligned}$$

By the definition of variance, $V[X] = E[X^2] - \left(E[X]\right)^2$

$$\begin{aligned}
V[X] &= E[X^2] - \left(E[X]\right)^2 \\
&= \sum_{x=1}^{3} x^2 P\left(X = x\right) - (2.05)^2 \\
&= 1^2 p_x(1) + 2^2 p_x(2) + 3^2 p_x(3) - 4.2025 \\
&= 1(0.30) + 4(0.35) + 9(0.35) - 4.2025 \\
&= 0.30 + 1.40 + 3.15 - 4.2025 \\
&= 4.85 - 4.2025 \\
&= 0.6475
\end{aligned}$$

Again, by the definition of variance, $V[Y] = E[Y^2] - \left(E[Y]\right)^2$

$$\begin{aligned}
V[Y] &= E[Y^2] - \left(E[Y]\right)^2 \\
&= \sum_{y=0}^{2} y^2 P\left(Y = y\right) - (1.15)^2 \\
&= 0^2\left(P\left(Y = 0\right)\right) + 1^2\left(P\left(Y = 1\right)\right) + 2^2\left(P\left(Y = 2\right)\right) - 1.3225 \\
&= 0\sum_{x=1}^{3} p(x,0) + 1\sum_{x=1}^{3} p(x,1) + 4\sum_{x=1}^{3} p(x,2) - 1.3225 \\
&= 0 + (0.10 + 0.20 + 0.05) + 4(0.05 + 0.10 + 0.25) - 1.3225 \\
&= 1.95 - 1.3225 \\
&= 0.6275
\end{aligned}$$

And to conclude,

$E[X] = 2.05$	$E[Y] = 1.15$
$V[X] = 0.6475$	$V[Y] = 0.6275$

(f) Find $\text{Cov}(X, Y)$.

Solution: Using the formula

$$\text{Cov}(X, Y) = E[XY] - \mu_x \mu_y$$

We simply need to calculate $E[XY]$ as we already have μ_x and μ_y from $E[X]$ and $E[Y]$ respectively. We can directly compute $E[XY]$ from definition.

$$\begin{aligned}
E[XY] &= \sum_{(x,y)\in S}\sum xy \cdot p(x,y) \\
&= \sum_{x=1}^{3}\sum_{y=0}^{2} xy \cdot p(x,y) \\
&= \sum_{x=1}^{3}\sum_{y=1}^{2} xy \cdot p(x,y) \\
&= p(1,1) + 2p(1,2) + 2p(2,1) + 4p(2,2) + 3p(3,1) + 6p(3,2) \\
&= 1(0.10) + 2(0.05) + 2(0.20) + 4(0.10) + 3(0.05) + 6(0.25) \\
&= 2.65
\end{aligned}$$

Substituting back into the formula,

$$\begin{aligned}
\text{Cov}(X,Y) &= E[XY] - \mu_x\mu_y \\
&= E[XY] - 2.05 \cdot 1.15 \\
&= 2.65 - 2.3575 \\
&\boxed{= 0.2925}
\end{aligned}$$

(g) Suppose that $U = 3X - 2Y$. Use Theorem 5.12 to find the mean and variance of U.

Solution: These can be computed by the linearity definitions. We'll start with mean,

$$\begin{aligned}
\mu_u &= E[U] \\
&= E[3X - 2Y] \\
&= E[3X] - E[2Y] \\
&= 3E[X] - 2E[Y] \\
&= 3(2.05) - 2(1.15) \\
&= 3.85
\end{aligned}$$

Next for variance, we will follow a similar method, using a modified Theorem 5.12,

$$V[aY_1 + bY_2] = a^2V[Y_1] + b^2V[Y_2] + 2ab\,\text{Cov}(X,Y)$$

$$\begin{aligned}
\sigma_u^2 &= V[U] \\
&= V[3X - 2Y] \\
&= V[3X] + V[-2Y] + 2ab\,\text{Cov}(Y_1,Y_2) \\
&= 3^2V[X](-2)^2V[Y] + 2(3)(-2)\,\text{Cov}(X,Y) \\
&= 9(0.6475) + 4(0.6275) - 12(0.2925) \\
&= 4.8275
\end{aligned}$$

$$\boxed{E[U] = 3.85 \qquad\qquad V[U] = 4.8275}$$

4. The time needed to complete a certain factory job is a normal random variable with mean $\mu = 50$ minutes and standard deviation $\sigma = 5$ minutes.

   (a) What is the probability that a (randomly selected) job will be completed in 53 minutes or less?

   Solution: We start by normalizing the distribution, so

$$Z = \frac{\bar{Y} - \mu}{\sigma}$$

   $\mu = 50$ is given, and $\sigma = 5$ is given, so we can directly substitute these into the Z-score equation. Since a success is defined as ≤ 53, then $\bar{Y} = 53$

$$Z = \frac{53 - 50}{5} = \frac{3}{5} = 0.6$$

   Since Table 4 gives upper tail, table_4(0.6) $\approx \int_{0.6}^{\infty} f$, so we need $1 -$ table_4(0.6), thus,

$$\begin{aligned} P(Y \leq 53) &= 1 - P(Y \geq 53) \\ &= 1 - \text{table\_4}(0.6) \\ &\approx 1 - 0.2743 \\ &\boxed{\approx 0.7257} \end{aligned}$$

   (b) What is the probability that the average time of ten randomly selected jobs will be less than 53 minutes or less?

   Solution: For $n = 10$, $\bar{M}$ is distributed by $N(50, 5^2/10)$, therefore $\sigma = \sqrt{5^2/10} = \sqrt{10}/2$. The computation follows

$$\begin{aligned} P(\bar{M} \leq 53) &= P\left(Z \leq \frac{53 - 50}{\sqrt{10}/2}\right) \\ &= P\left(Z \leq \frac{3}{\sqrt{10}/2}\right) \\ &= P\left(Z \leq \frac{3\sqrt{10}}{5}\right) \\ &\approx P\left(Z \leq 1.8973666\right) \\ &\approx 1 - \text{table\_4}(1.90) \\ &\approx 1 - 0.0287 \\ &\boxed{\approx 0.9713} \end{aligned}$$

(c) Only 5% of the time will a single job be completed in M minutes or less. Determine M.

Solution: Let λ represent some Z-score. Then

$$P(Y \leq M) = 0.05 = 1 - \text{table\_4}(\lambda)$$

In order for the equality to hold, we need $\text{table\_4}(\lambda) = 0.95$, but $\text{table\_4}(x)$ is defined for $x \geq 0$. So we need to instead find a $\text{table\_4}(-\lambda) = 0.05$ by symmetry, and thus $-\lambda \approx 1.645 \implies \lambda \approx -1.645$

We now need

$$Z = \lambda \approx -1.645 \approx \frac{\bar{Y} - \mu}{\sigma} = \frac{\bar{Y} - 50}{5}$$

Which implies,

$$5(-1.645) \approx \bar{Y} - 50$$

$$-8.225 \approx \bar{Y} - 50$$

$$\bar{Y} \approx 50 - 8.225$$

$$\bar{Y} \approx 41.775$$

$$\boxed{\text{Thus, } M \approx 41.775 \text{ minutes}}$$