Worksheet 1

Q1

Compute the partial derivatives of $f(x,y) = x\ln(y)$ with respect to (x,y). The partial derivatives are simply the derivatives for each variable, assuming the others are fixed. Computing each partial derivative uses the exact same rules as derivatives for single variables.

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix}$$

Answer:

$$\nabla f(x,y) = \begin{bmatrix} \ln(y) \\ \frac{x}{y} \end{bmatrix}$$

Q2

Compute the Jacobian matrix of f with respect to (x,y,z),

$$f(x,y,z) = \begin{bmatrix} f_1(x,y,z) \\ f_2(x,y,z) \\ f_3(x,y,z) \end{bmatrix} = \begin{bmatrix} xyz \\ y^2 \\ x+z \end{bmatrix}.$$

Answer:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1(x,y,z)}{\partial x} & \frac{\partial f_1(x,y,z)}{\partial y} & \frac{\partial f_1(x,y,z)}{\partial z} \\ \frac{\partial f_2(x,y,z)}{\partial x} & \frac{\partial f_2(x,y,z)}{\partial y} & \frac{\partial f_2(x,y,z)}{\partial z} \\ \frac{\partial f_3(x,y,z)}{\partial x} & \frac{\partial f_3(x,y,z)}{\partial y} & \frac{\partial f_3(x,y,z)}{\partial z} \end{bmatrix} = \begin{bmatrix} yz & xz & xy \\ 0 & 2y & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Q3

Suppose that in a lottery you have 0.01% chance of winning and the prize is $1000. The ticket to enter the lottery costs you $10. What is the expected amount you would earn, when buying a ticket for this lottery?

Answer:

$$\mathbb{E}[X] = \left(\frac{0.01}{100}\right)(1000 - 10) + \left(1 - \frac{0.01}{100}\right)(0 - 10) = -9.9$$

Q4

Adam and Martha propose a simple dice game to you. You can throw a die up to two times, and they will reward you with the amount equivalent to the face value of the die. If you throw a die once and 3 comes up, you can choose to take $3 or throw again. If you choose to throw again and 2 comes up, you earn only $2. The amount you earn is not additive and you only earn the amount of your last roll.

1. Suppose in your first roll, the dice comes up as a 1. What is the expected amount you would earn in your second roll?

2. For what values in your first roll should you re-roll the die?

3. What is the expected amount you would earn in this game if you play optimally?

Answer:

1. Assume a fair dice. The first die roll will not affect the second roll in any manner. So the expectation of the amount you would earn in the first roll is independent of the outcome of the first die, and can be calculated separately. If we let X denote the random variable representing the amount earned in the second throw, then

$$\mathbb{E}[X] = \frac{1}{6} \times (1 + 2 + 3 + 4 + 5 + 6) = 3.5.$$

2. We should re–roll the die in case we get 1, 2, or 3. Since, re–rolling it will give an expected sum of 3.5.

3. The optimal policy would be to roll the dice if we get 1, 2, or 3, and not roll it again if we get 4, 5, or 6. In this case the expected gain would be simply

$$\underbrace{3 \times \frac{1}{6} \times \frac{1}{6} \times (1 + 2 + 3 + 4 + 5 + 6)}_{\text{Rolling the dice again incase of 1, 2, or 3}} + \underbrace{\frac{1}{6} \times (4 + 5 + 6)}_{\text{Not rolling the dice again}} = 4.25.$$

Q5

Prove the tower property
$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}[X|Y]\right].$$

Answer:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_x x \cdot \Pr(x) \\
&= \sum_x x \sum_y \Pr(x|y) \cdot \Pr(y) \\
&= \sum_y \left(\sum_x x \cdot \Pr(x|y) \right) \cdot \Pr(y) \\
&= \sum_y \mathbb{E}[X|Y = y] \cdot \Pr(y) \\
&= \mathbb{E}\left[\mathbb{E}[X|Y]\right]
\end{aligned}
$$

Q6

In real world problems, we rarely know the distribution of a variable. Instead, we often make assumptions on its distribution. One commonly assumed distribution for continuous random variables is the Gaussian (or Normal) distribution. List at least two reasons that justifies using the Gaussian distribution in practice.

Answer:

1. Because of Central Limit Theorem (distribution of sample mean, i.e. mean of random variables drawn from arbitrary distributions, approaches a Normal distribution as the number of samples increases). If we assume that a real world process has many underlying contributing factors (coming from arbitrary distributions), then a Normal distribution is likely to model it well.

2. Because of simplicity of the Normal distribution: it just has two parameters μ and σ, and has nice properties, for example it is a conjugate distribution.

Q7

Given $\{y_1, y_2, \ldots, y_{n-1}, y_n\}$ and $f(y) = (\theta + 1)y^\theta, 0 < y < 1.$, find the MLE for θ.

Answer:

The Maximum Likelihood Solution is defined by

$$\theta_{MLE} := \arg\max_\theta \ln\left(\prod_{i=1}^n f(y_i)\right) = \arg\max_\theta \sum_{i=1}^n \ln\left((\theta+1)y_i^\theta\right).$$

We solve this by taking the derivative of the term following arg max with respect to θ and setting it to zero.

$$\frac{\partial}{\partial\theta}\sum_{i=1}^n \ln\left((\theta+1)y_i^\theta\right) = \sum_{i=1}^n \frac{\partial}{\partial\theta}\left[\ln(\theta+1) + \theta\ln y_i\right] = \sum_{i=1}^n \left[\frac{1}{\theta+1} + \ln y_i\right] = 0$$

or
$$\frac{n}{\theta+1} = -\sum_{i=1}^n \ln y_i$$

or
$$\theta_{MLE} = -\frac{n}{\sum_{i=1}^n \ln y_i} - 1.$$

Q8

What do the 3rd and 4th central moments represent respectively (knowing that the 1st raw moment is the mean and the 2nd central moment is the variance).

Answer:

3rd moment is Skewness (a measure of symmetry of a distribution), and the 4th moment is Kurtosis (a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution). See: `https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm`

Q9

Consider the following game: you roll two (fair) 6-sided dice and win \$1 if the sum of the dice roll is 2, 5, 7, 8 or 11. Otherwise, you lose \$1.

1. What is the expected value of the sum of the two dice?

2. What is the variance of the sum of the two dice?

3. What is the expected value of the winnings for playing this game. In other words, how much money are you expected to gain (or lose).

4. What is the variance of the winnings for this game?

5. Would you play this game as stated above? How about if the amount won or lost was \$100? How about \$1000?

Answer:

1. Let X_1 and X_2 denote the random variables representing the value on the first and second roll of the dice respectively.

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 2\,\mathbb{E}[X_1] = 2 \times \frac{1}{6} \times (1 + 2 + 3 + 4 + 5 + 6) = 2 \times 3.5 = 7.$$

2. For instructive purpose let us analyze the variance of the sum of the two dice first:

$$\begin{aligned}
\mathrm{Var}[X_1 + X_2] &:= \mathbb{E}\left[\left(X_1 + X_2 - \mathbb{E}[X_1 + X_2]\right)^2\right] \\
&= \mathbb{E}[(X_1 + X_2)^2] - \mathbb{E}[X_1 + X_2]^2 \\
&= \mathbb{E}[X_1^2 + X_2^2 + 2X_1X_2] - \mathbb{E}[X_1]^2 - \mathbb{E}[X_2]^2 - 2\,\mathbb{E}[X_1]\,\mathbb{E}[X_2] \\
&= \left(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\right) + \left(\mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2\right) + 2\left(\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]\right) \\
&= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\mathrm{Cov}(X_1, X_2) \\
&= \mathrm{Var}[X_1] + \mathrm{Var}[X_2]. \qquad\qquad \text{(since } X_1 \text{ and } X_2 \text{ are independent)}
\end{aligned}$$

Now continuing a similar analysis we did before, we get

$$\begin{aligned}
\mathrm{Var}[X_1 + X_2] &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] = 2\mathrm{Var}[X_1] \\
&= 2\left(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\right) \\
&= 2 \times \left(\frac{1}{6} \times (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3.5^2\right) \\
&= 5.83.
\end{aligned}$$

3. We have the following combinations for the digits on the dice for which we gain \$1:

   - $X_1 + X_2 = 2:$ $\{(1,1)\}$,
   - $X_1 + X_2 = 5:$ $\{(1,4), (2,3), (3,2), (4,1)\}$,
   - $X_1 + X_2 = 7:$ $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$,
   - $X_1 + X_2 = 8:$ $\{(2,6), (3,5), (4,4), (5,3), (6,2)\}$, and
   - $X_1 + X_2 = 11:$ $\{(5,6), (6,5)\}$.

   With this information, the expectation is straightforward to calculate. If we let W represent the random variable denoting the amount of winnings:

$$\mathbb{E}[W] = \frac{18}{36} \times (+1) + \frac{18}{36} \times (-1) = 0.$$

4. Using the above information we get

$$\mathrm{Var}[W] = \mathbb{E}[W^2] - \mathbb{E}[W]^2 = \left(\frac{18}{36} \times (+1)^2 + \frac{18}{36} \times (-1)^2\right) - 0^2 = 1.$$

5. It is equally likely for us to win or lose the money in this game – so it is purely dependent on luck. It would be more wise to play this game for a smaller amount if we want to minimize the risk of losing money. On the flip side, playing with a smaller amount would also decrease the potential gains.

Q10

Now suppose a game where you choose to flip one of two (possibly unfair) coins. You win $1 if your chosen coin shows heads and lose $1 if it shows tails. Note that you do not know the probability of the coin outcomes. Instead, you are able to view 6 sample flips for each coin respectively: (T,H,H,T,T,T) and (H,T,H,H,H,T).

1. For each coin: what is the MLE estimate for the probability of heads p? Construct a confidence interval for p.

2. Which coin would you flip? Would you be willing to flip a coin other than the one you chose?

Answer:

1. Represent T by 0 and H by a 1. Then the MLE estimate of the probability of getting a heads is just the mean of the input data. If p_1 and p_2 denote the probability of heads for coin 1 and coin 2 respectively, then $p_1 = \frac{2}{6} = 0.33$ and $p_2 = \frac{4}{6} = 0.67$.

   To compute the confidence intervals, we will first need to compute the sample standard error (standard deviation over the square root of sample size) σ_1^- and σ_2^- for these two coins. The standard deviations for the two coins are $\sigma_1 = \sqrt{\frac{2}{6} \times 1 - \left(\frac{2}{6}\right)^2} = 0.471$ and $\sigma_2 = \sqrt{\frac{4}{6} \times 1 - \left(\frac{4}{6}\right)^2} = 0.471$. Then $\sigma_1^- = \frac{\sigma_1}{\sqrt{6}} = 0.192$ and $\sigma_2^- = \frac{\sigma_2}{\sqrt{6}} = 0.192$.

   Assuming the underlying sample mean distribution to be Normal (central limit theorem), the 95% confidence interval then are $\mu \pm 1.96\sigma^-$. We have $1.96\sigma^- = 0.38$. We thus get,

   $$-0.05 \leq p_1 \leq 0.71 \text{ and } 0.29 \leq p_2 \leq 1.05 \text{ with } 95\% \text{ confidence.}$$

2. Its better to flip the second coin based on the available information, since getting a heads is more favourable in our case.

   We might be willing to flip the first coin as well, especially because the our confidence in the estimates of p_1 and p_2 are not very high. It might be possible that $p_1 > p_2$. We could be more sure with a larger number of flip trials of both the coins.

Q11

Suppose you start a betting game with $5000 and roll a 6 sided die 100 times. Before any dice rolls however, you must decide on a proportion of your wealth to be wagered for each dice roll. If the die shows 1 or 2, you will lose 12% of your wager. If the die shows 3, 4, 5 or 6 you will win 10% of your wager.

1. What is a good proportion to bet?

2. How much are you expected to win with this proportion?

3. What is the probability of you losing over 50% of your wealth by the end of the game?

4. What is the probability of you losing any money at all?

Answer:

Q12

We say an estimator is unbiased if its expected value equals the true value of the parameter being estimated. For example, if we roll a die 10 times and denote the value of the ith roll to be X_i, then the sample mean (average roll) is an unbiased estimator of the true mean since

$$E\left[\frac{\sum_{i=1}^{10} X_i}{10}\right] = E[X].$$

Where $E[X]$ is the expected value of a roll.

Suppose you would like to estimate what the expected roll of a fair die is but you only have a loaded die with probability mass function $\tilde{p}(x)$. Show that $Y = X\frac{1/6}{\tilde{p}(X)}$ is an unbiased estimator of the mean for a roll with a fair die, where X is the outcome of the loaded die.

Answer:

$$
\begin{aligned}
\mathbb{E}_{\text{loaded die}}\left[X\frac{1/6}{\tilde{p}(x)}\right] &= \sum_{x\in\{1,2,3,4,5,6\}} \tilde{p}(x) \cdot \left(X\frac{1/6}{\tilde{p}(x)}\right) \\
&= \sum_{x\in\{1,2,3,4,5,6\}} X \cdot \frac{1}{6} \\
&= \mathbb{E}_{\text{fair die}}\left[X\right].
\end{aligned}
$$

The ratio $\frac{1/6}{\tilde{p}(x)}$ is known as the importance sampling ratio. We may discuss more about it later in the course, and its connection to Off–policy learning.