DS-GA 1002 HW 7

1. (a) Yes, Let's assume that random variable a takes values of negative, zero, positive integer (ex) $\{-2, 0, 2\}$), with equal probability.

$$(E(\tilde{a}))^2 = \left(\frac{-2+0+2}{3}\right)^2 = 0^2, \text{ while } E(\tilde{a}^2) = \frac{4+0+4}{3} = \frac{8}{3},$$

So, $E^2(\tilde{a}) < E(\tilde{a}^2)$. That is $E^2(\tilde{a})$ can be less than $E(\tilde{a}^2)$.

(b) True, $E(\tilde{a}) = E(\tilde{b})$ since $\tilde{a}$ and $\tilde{b}$ have the same distribution.
(E operator only depends on pmf(pdf))
By independence, $E(\tilde{a}\tilde{b}) = E(\tilde{a}) \cdot E(\tilde{b}) = E^2(\tilde{a})$

(c) Probability of event occurring is $1/n$.

So, $\begin{cases} P_{I_i}(0) = 1 - \frac{1}{n}, \\ P_{I_i}(1) = \frac{1}{n} \end{cases}$

By linearity of expectation
E(number of kids that get their own present)

$$= E\left(\sum_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} E(I_i) = \sum_{i=1}^{n} P_{I_i}(1) = \sum_{i=1}^{n} \frac{1}{n} = \frac{1}{n} \times n = 1$$

∴ Expected number of children that end up getting the present bought by their own parents is 1.

2.(a) $E(\tilde{t}|\tilde{x})$ is a function of $\tilde{x}$ (so determined by x).

Since $\tilde{x}$ is discrete, $E(\tilde{t}|\tilde{x})$ is a discrete random variable.

So its pmf is as below:

Let $\tilde{x} = h(\tilde{a})$ where $h(a) = E(\tilde{t}|\tilde{x}=d)$,

since $E(\tilde{t}|\tilde{x}=d)$ can be either $E(\tilde{t}|\tilde{x}=1)=\frac{1}{2}$ or $E(\tilde{t}|\tilde{x}=0)=1$,

$$\boxed{\text{Pmf}: \begin{cases} P_{\tilde{x}}(\frac{1}{2}) = p(\tilde{x}=1) = 0.1 \\ P_{\tilde{x}}(1) = p(\tilde{x}=0) = 0.9 \end{cases}}$$

(b) For exponential distribution, if parameter is λ,

mean is $\frac{1}{\lambda}$, and mean square is $2/\lambda^2$.

$$Var(\tilde{t}) = E(\tilde{t}^2) - \{E(\tilde{t})\}^2$$

$$\left(\begin{aligned} E(\tilde{t}) &= \tfrac{1}{2} \times 0.1 + 1 \times 0.9 = 0.95 \\ E(\tilde{t}^2) &= 2/_{2^2} \times 0.1 + 2/_{1^2} \times 0.9 = 1.85 \end{aligned} \right)$$

$$= 1.85 - (0.95)^2 = \boxed{0.9475.}$$

2-(c) Let's assume $\tilde{c}$ as a computer that will either break down or not during the first year.

Then, we can say that $\tilde{c}$ is a binomial random variable with parameter n and θ.

$$\begin{cases} n = 100, \\ \theta = \int_0^1 p(\tilde{\alpha}=0)\, f_{\tilde{t}|\tilde{\alpha}}(\tilde{t}|\tilde{\alpha}=0)\,dt + \int_0^1 p(\tilde{\alpha}=1)\, f_{\tilde{t}|\tilde{\alpha}}(\tilde{t}|\tilde{\alpha}=1)\,dt \end{cases}$$

$$\left(\text{Since, } \begin{cases} f(\tilde{t}|\tilde{\alpha}=1) = 2 \cdot e^{-2t} \\ f(\tilde{t}|\tilde{\alpha}=0) = e^{-t} \end{cases}, \quad \begin{cases} p(\tilde{\alpha}=1) = 0.1 \\ p(\tilde{\alpha}=0) = 0.9 \end{cases} \right)$$

$$= \int_0^1 0.9 e^{-t}\,dt + \int_0^1 0.1 \times 2 \times e^{-2t}\,dt$$

$$= 0.9 \times \left[-e^{-t}\right]_0^1 + 0.2 \times \left[-\tfrac{1}{2}e^{-2t}\right]_0^1$$

$$= 0.9 \times (-e^{-1}+1) + 0.2 \times \left[-\tfrac{1}{2}e^{-2} + \tfrac{1}{2}\right)$$

$$\fallingdotseq 0.655.$$

$$\boxed{\begin{aligned} \therefore\ E(\tilde{c}) &= n\theta = 100 \times 0.655 \fallingdotseq 65.5 \\ Var(\tilde{c}) &= n\theta(1-\theta) = 100 \times 0.655 \times (1-0.655) \fallingdotseq 22.5975 \end{aligned}}$$

3.(a) It is a number. Since $\tilde{a}$ is constant ($\tilde{a}=a$),

and if we say its value is a fixed value x, $Var(\tilde{b}/\tilde{a}=a)$ is a number.

It represents the variance of the random variable $\tilde{b}$ given the information $\tilde{a}=a$. Also, it can be interpreted as a variance of a new distribution of the random variable $\tilde{b}/(\tilde{a}=a)$

$$\left(Var(\tilde{b}/\tilde{a}=a) = E\left((\tilde{b} - E(\tilde{b}/\tilde{a}=a))^2 \mid \tilde{a}=a \right) \right)$$

(b) Apart from 3-(a), we now treat x as a variable.
For any x we assign a value $Var(\tilde{b}/\tilde{a}=a)$,
this is a function from the range of $\tilde{a}$ to real numbers,
which we call by h, so $h(x) = Var(\tilde{b}/\tilde{a}=a)$.
Since this function is defined on a probability space,
we can regard it as a random variable by $h(\tilde{a}) = Var(\tilde{b}/\tilde{a})$

(c) Using iterated expectations,

$$E(Var(\tilde{b}/\tilde{a})) = E(E(\tilde{b}^2/\tilde{a})) - E(E(\tilde{b}/\tilde{a})^2)$$
$$= E(\tilde{b}^2) - E(E(\tilde{b}/\tilde{a})^2)$$
$$Var(E(\tilde{b}/\tilde{a})) = E(E(\tilde{b}/\tilde{a})^2) - E(E(\tilde{b}/\tilde{a}))^2$$
$$= E(E(\tilde{b}/\tilde{a})^2) - E(\tilde{b})^2$$
$$\rightsquigarrow E(Var(\tilde{b}/\tilde{a})) + Var(E(\tilde{b}/\tilde{a})) = E(\tilde{b}^2) - \{E(\tilde{b})\}^2 = Var(\tilde{b})$$

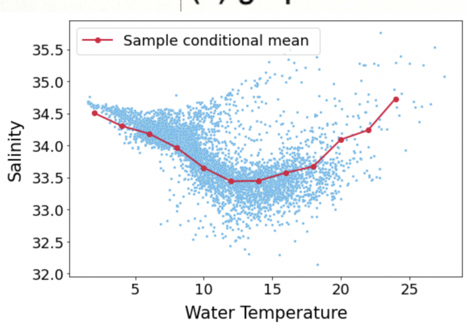
Variance $\tilde{b}$ is an average of the variance of $\tilde{b}$ given $\tilde{a}$,
plus the variance of the average of $\tilde{b}$ given $\tilde{a}$.

4-a) According to the conditional means, the salinity decreases with temperature and then increases again.

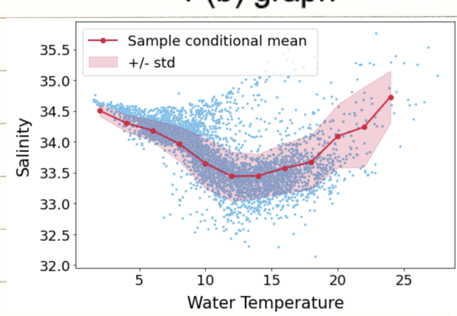
```
max_val = np.max(temp)
width_bin = 2
fig = plt.figure(figsize = (9,6))
plt.scatter(temp,salnty, s=5, c="dodgerblue", marker='o', edgecolor="skyblue")

# TODO: create bins from 0 to the maximum to discretize continous temperatures
### INSERT CODE HERE ###
grid = np.arange(0, max_val, width_bin)

# TODO: Compute the conditional expection of sanlinity given temperture
cond_average_salnty = np.zeros(len(grid))

### INSERT CODE HERE ###
for i in range(len(grid)):
    if i == len(grid)-1:
        cond_average_salnty[i] = np.mean(salnty[np.where(temp>=grid[i])])
        break
    cond_average_salnty[i] = np.mean(salnty[np.where((temp>=grid[i]) & (temp<=grid[i+1]))])

plt.plot(grid[1:-1],cond_average_salnty[1:-1],'-o',lw=2,color='crimson', label="Sample conditional mean ")
plt.ylabel("Salinity", fontsize=21,labelpad=10)
plt.xlabel("Water Temperature", fontsize=21,labelpad=10)
```

```
# 4-b)
# TODO: Compute the conditional standard deviation of sanlinity given temperture
cond_std_salnty = np.zeros(len(grid))

### INSERT CODE HERE ###
for i in range(len(grid)):
    if i == len(grid)-1:
        cond_std_salnty[i] = np.std(salnty[np.where(temp>=grid[i])])
        break
    cond_std_salnty[i] = np.std(salnty[np.where((temp>=grid[i]) & (temp<=grid[i+1]))])

fig = plt.figure(figsize = (9,6))
plt.scatter(temp,salnty, s=5, c="dodgerblue", marker='o', edgecolor="skyblue")


plt.plot(grid[1:-1],cond_average_salnty[1:-1],'-o',lw=2,color='crimson', label="Sample conditional mean ")
plt.fill_between(grid[1:-1], cond_average_salnty[1:-1]-cond_std_salnty[1:-1],
                 cond_average_salnty[1:-1]+cond_std_salnty[1:-1], color='crimson', alpha=0.2, label="+/- std")

plt.ylabel("Salinity", fontsize=21,labelpad=10)
plt.xlabel("Water Temperature", fontsize=21,labelpad=10)
```

4-(a) graph



4-(b) graph



(c) The reliability of the conditional mean estimator depends on the number of data that fell into the bin. As you can see in the above graph, the conditional mean estimates to the left is more reliable than that of right, as each bin have more data points.