Homework 5 **Name: Jian Wang** **FSUID:JW09R**

Problem 1

Answer:

The formula of the tilted mass function is as follows:

$$f_t(x) = \frac{e^{tx} f(x)}{M(t)}$$

Where $M(t) = E[e^{tx}]$.

In this case: x will be 0 at the probability p and be 0 at the probability (1-p)

So the M(t) = (1-p)+pe^{t}

Then we can find the

$f_t(x) = \frac{e^{tx} f(x)}{M(t)} = \frac{(Pe^t)^x (1-p)^{1-x}}{(1-p)+Pe^t}$ here x=0 or 1

We can see from the above formula, $f_t(1) = \frac{(Pe^t)}{(1-p)+Pe^t}$ when x= 1

And $f_t(0) = \frac{(1-p)}{(1-p)+Pe^t}$ when x= 0

The sum of the $f_t(1)$ and $f_t(0)$ is 1 , so we can conclude that the tilted mass function follows a bernouli distribution with p= $\frac{(Pe^t)}{(1-p)+Pe^t}$ and $q = \frac{(1-p)}{(1-p)+Pe^t}$

Problem 2:

Answer(a) : From the definition of the expectation, we can know that

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-\infty}^{+\infty} x * x * f(x)dx \leq \int_{-\infty}^{+\infty} a * x * f(x)dx \text{ (since } 0 \leq x \leq a)$$

$= a * \int_{-\infty}^{+\infty} x * f(x)dx = aE(x)$

so $E(x^2) \leq aE(x)$

Answer(b): Var(x) = $E(x^2) - (E(x))^2 \leq aE(x) - (E(x))^2 \text{(from a)} = E(x)(a - E(x))$

Answer (c) : From the inequality, we know that $\sqrt{ab} \leq \frac{a+b}{2}$ when $a, b \geq 0$

Since both $E(x)$ and $(a - E(x))$ greater than 0

Then $\sqrt{E(x) * (a - E(x)} \leq \frac{E(x)+a-E(x)}{2} = \frac{a}{2}$

Then $\sqrt{E(x) * (a - E(x)} \leq \frac{a^2}{4}$

Form the question (b); we can derived that Var(x) = $E(x^2) - \left(E(x)\right)^2 \leq E(x)\left(a - E(x)\right) \leq \frac{a^2}{4}$

The proof is completed.

Problem 3:

Answer:
Firstly, since we know that time can't inverse, so when i<j then $t_i < t_j$ therefore, when i<j then min(t_i, t_j)=t_i
Secondly, in the time status matrix $[t_{i,j]}]$
For each i, it can be appeared in the i row or i column, so the max numbers t_i can appear is 2*n-1, the reason for minus 1 is that there will be an overlap in the i row and j column. Besides for each i when the status of factor in both the row and the column is less than i, then its value of T will less than t_i. Since it will occurred in both the row and the column, then we should minus 2*(i-1) from the 2*n-1.

So the total number for i is 2*n-1-2*(i-1)=2n-(2i-1)

Therefore, we can get that $\sum_{i,j} \min\left(t_i, t_j\right) = \sum_{i=1}^{n}\left(2 * n - (2i - 1)\right)t_i$

I also give an example when i,j=5

$$\begin{array}{ccccc} t_1 & t_1 & t_1 & t_1 & t_1 \\ t_1 & t_2 & t_2 & t_2 & t_2 \\ t_1 & t_2 & t_3 & t_3 & t_3 \\ t_1 & t_2 & t_3 & t_4 & t_4 \\ t_1 & t_2 & t_3 & t_4 & t_5 \end{array}$$

We can see the number of t_1 is 9 = 2*5-(2*1-1)
the number of t_2 is 7 = 2*5-(2*2-1)
the number of t_3 is 5 = 2*5-(2*3-1)
the number of t_4 is 3 = 2*5-(2*4-1)
the number of t_5 is 1 = 2*5-(2*5-1)

which matches the identity.

Problem 4:

Answer:

We can see that Var($\beta Y + (1 - \beta)C) = \beta^2 Var(Y) + (1 - \beta)^2 Var(c) + 2\beta(1 - \beta)Cov(Y, C)$

The first order is 2* $\beta * Var(Y) - 2 * Var(C) + 2 * \beta * Var(c) + 2 * Cov(Y, C) - 4 * \beta * Cov(Y, C)$

We take the first order equals to zero and get the $\beta^*=\frac{\mathrm{var}\,(c)-\mathrm{Cov}\,(Y,C)}{\mathrm{Var}\,(Y)+\mathrm{Var}\,(C)-2\mathrm{Cov}\,(Y,C)}$

The second order is $2*\mathrm{Var}(Y)+2*\mathrm{Var}(C)-4*\mathrm{Cov}(Y,C)$=2Var(Y-C) is always greater than 0.
So the β^* can derive the minium variance.

If C=E(Y|Z)
Then Var(C)= $E(c^2)$-$(E(C))^2$ we know that E(c)=E(Y)

And E$(c^2)=E\big(E(Y|Z)*E(Y|Z)\big)=E(E(Y*E(Y|Z)|Z)(\text{since } E(Y|Z)\text{is Z measurable})$

So E$(c^2)=E\big(E(YC|Z)\big)=E(YC)$

Then we can get Var(c)=E$(C^2)-(E(C))^2$= E(YC)-$(E(Y))^2=E(YC)-E(Y)*E(C)$=Cov(Y,C)

From the formula $\beta^*=\frac{\mathrm{var}\,(c)-\mathrm{Cov}\,(Y,C)}{\mathrm{Var}\,(Y)+\mathrm{Var}\,(C)-2\mathrm{Cov}\,(Y,C)}$

We can see that when C=E(Y|Z) then $\beta^*=0$ which means that no further improvement is possible by combining Y and E(Y|Z)

Problem 5:

Answer (a):
1)
The sample mean estimator for θ is

$$\overline{X}=\frac{X_1+X_2+,\dots,+X_n}{n}$$

Where E(X)= θ

We have
E($\overline{X}$)=E(X)= θ
Var($\overline{X}$) $=\sigma_x{}^2/n$

In this case:
Let Y=e^x where x follows the U(0,1) distribution

Then E(Y)=$\int_0^1 e^x dx=e-1$ which equals to θ

$$E(Y^2)=\int_0^1 e^{2x}dx=0.5(e^2-1)$$

So the Var (Y) is equal to 0.5$(e^2-1)-(e-1)^2=2e-0.5e^2-1.5$=0.242036

The sample mean estimator is

$\overline{Y} = \frac{Y_1 + Y_2 +, \dots, +Y_n}{n}$

Var($\overline{Y}$) $= {\sigma_y}^2/n$ =0.242036/n

2) The antithetic variates estimator is as follows:

Z=1/2*(X+Y) where X and Y is an unbiased estimator for θ

In this case, we let Y equals to e^x and Z equals to e^{1-x} here, x follows the U(0,1)

The antithetic variates estimator is

W=1/2*(Y+Z)

Var(W)=1/4*Var(Y)+1/4*Var(Z)+1/2*cov(Y,Z)

Var (Y) =Var(Z) $= 2e - 0.5e^2 - 1.5$

E(YZ)=$\int_0^1 e^x * e^{1-x} dx = e$

So cov(Y,Z)= E(YZ)-E(Y)E(Z)=e-$(e-1)^2$=3*e-e^2-1<0

Then the Var(W) = e-0.25$e^2 - 0.75$+1.5e-0.5$e^2 - 0.5$=2.5e-0.75$e^2 - 1.25$=0.003912

So the Var(Z)=1/n*var(W)=0.003912/n

From the above calculation, we can know that 0.003912/0.242036=0.016165, which means 98.4% of variance has been reduced

Answer(b):

The method of the control variates estimator is as follows:

$Y(\beta) = Y - \beta(C - u_c)$

Here the u_c is the mean value of C

In this case， we use the f(U)=U as the control variate, so the mean is 0.5

And the var(c) is 1/12

And the var($Y(\beta)$) is Var(Y)-2*β $* \mathrm{cov}(Y, C) + \beta^2 \mathrm{Var}(c)$

E(YC)=$\int_0^1 x * e^x dx = 1$ so the cov(Y,C)=1-0.5*(e-1) =1.5-0.5*e

We also used the $\beta^* = \frac{\mathrm{cov}\,(Y,C)}{\mathrm{Var}\,(C)}$ =(1.5-0.5e)*12

So var($Y(\beta)$)= $2e - 0.5e^2 - 1.5$-2*12*(1.5-0.5e)^2+12*(1.5-0.5e)^2=0.00394

SO the Var(Z)= var($Y(\beta)$)/n=0.00394/n

Compared with the variance in (a), we can know that the control variates estimator in this case is better than the sample mean estimator but worse than antithetic variates method.

We can see that it is less than the value of the other two method.

Answer(C):

I used the Randu method(Which is known not a good random generator) to generate the random numbers and calculate the expectation under the following three method:
For the sample mean estimator:

Us e the $\overline{\mathrm{X}} = \frac{\mathrm{X1+,...+Xn}}{\mathrm{N}}$

For the control variate estimator:

Use: $\overline{\mathrm{Y}} = \frac{\mathrm{X1+,...+Xn}}{\mathrm{N}} + \beta^{*}(\frac{\mathrm{C1+,...+Cn}}{\mathrm{N}} - \mu_{c})$

For the antithetic variates estimator
Use

$$\overline{\mathrm{X}} = \frac{1}{2\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{g(ui)} + \mathrm{g}(1-\mathrm{ui})$$

Use N equals to 10000,20000,30000,40000,50000 and compare the results and the true value is e-1

N	Sample mean Estimator		antithetic variates estimator		control variates estimator	
	result	absolute error	result	absolute error	result	absolute error
10000	1.718179	-0.000103	1.717906	-0.000376	1.718454	0.000172
20000	1.720527	0.002245	1.717833	-0.000449	1.723231	0.004949
30000	1.723738	0.005456	1.718201	-0.000081	1.729308	0.011026
40000	1.723344	0.005062	1.718485	0.000203	1.728225	0.009943
50000	1.722052	0.003770	1.718483	0.000201	1.725631	0.007349

From the above result, we can see that the result of the antithetic method is the best.