

# **Probability & Statistics for EECS: Homework #12**

Due on May 7, 2023 at 23:59

Name: **Zhou Shouchen**  
Student ID: 2021533042

## Problem 1

(a) Since the prior distribution of  $p$  is  $Unif(0, 1)$ , which is exactly same as  $Beta(1, 1)$ .

This is because  $\beta(1, 1) = \int_0^1 x^{1-1}(1-x)^{1-1}dx = 1, x \in (0, 1)$ , then  $\frac{1}{\beta(1, 1)}x^{1-1}(1-x)^{1-1} = 1, x \in (0, 1)$ .

So the prior distribution is that  $p \sim Beta(1, 1)$ .

From what we have learned about the Bayesian Inference,

we can calculate the posterior distribution of  $p$ :

Since  $X_i|p \sim Bern(p)$ , so  $S = \sum_{i=1}^n X_i|p \sim Bin(n, p)$ ,

with Beta-Binomial conjugate prior, we can get the posterior is that:

$$(p|X_1 = x_1, \dots, X_n = x_n) = (p|S = k) \sim Beta(1 + \sum_{i=1}^n x_i, 1 + n - \sum_{i=1}^n x_i).$$

So above all, the posterior distribution of  $p$  given  $X_1 = x_1, \dots, X_n = x_n$  is that:

$$p|X_1 = x_1, \dots, X_n = x_n \sim Beta(1 + \sum_{i=1}^n x_i, 1 + n - \sum_{i=1}^n x_i).$$

And we can see that it only depends on the sum of  $x_i$ .

(b) With LOTP, we can get that

$$P(X_{n+1} = 1|X_1 + \dots + X_n = k) = \int_0^1 P(X_{n+1} = 1|X_1 + \dots + X_n = k, p = p')f_p(p'|X_1 + \dots + X_n = k)dp'$$

From (a), we know that  $p|X_1 = x_1, \dots, X_n = x_n \sim Beta(1 + k, 1 + n - k)$ ,

$$\text{so } f_p(p'|X_1 + \dots + X_n = k) = \frac{1}{\beta(1 + k, 1 + n - k)}p'^{1+k-1}(1-p')^{1+n-k-1} = \frac{1}{\beta(1 + k, 1 + n - k)}p'^k(1-p')^{n-k}.$$

So the above equation can be written as:

$$= \int_0^1 P(X_{n+1} = 1|X_1 + \dots + X_n = k, p = p') \frac{1}{\beta(1 + k, 1 + n - k)}p'^k(1-p')^{n-k}dp'$$

And since  $X_i|p$  are conditional independent  $Bin(p)$ ,

$$\text{so } P(X_{n+1} = 1|X_1 + \dots + X_n = k, p = p') = P(X_{n+1} = 1|p = p') = p',$$

So the above equation can be written as:

$$= \frac{1}{\beta(1 + k, 1 + n - k)} \int_0^1 p'p'^k(1-p')^{n-k}dp'$$

Since  $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ , and when  $a, b$  are positive integers,  $\Gamma(a) = (a-1)!$ , so  $\beta(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$ .

From the problem's description, we can get that both  $1 + k$  and  $1 + n - k$  are positive integers, so

$$\beta(1 + k, 1 + n - k) = \frac{k!(n-k)!}{(n+1)!}.$$

Also, we can see that  $\int_0^1 p'p'^k(1-p')^{n-k}dp' = \int_0^1 p'^{k+1}(1-p')^{n-k}dp' = \beta(k+2, n-k+1)$ .

From above analysis, we can also get that  $k+2, n-k+1$  are positive integers.

$$\text{So } \beta(k+2, n-k+1) = \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(k+2+n-k+1)} = \frac{(k+1)!(n-k)!}{(n+2)!}.$$

So the above equation can be written as:

$$= \frac{1}{\frac{k!(n-k)!}{(n+1)!}} \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k+1}{n+2}$$

So above all,  $P(X_{n+1} = 1|X_1 + \dots + X_n = k) = \frac{k+1}{n+2}$ .

(c) Reinterpret the Laplace's law of succession from the perspective of Beta-Binomial conjugacy is that: From the view of Bayesian Inference, we get a prior distribution of  $p$  is that  $p \sim Unif(0, 1) \sim Beta(1, 1)$ . And the  $n$  variables  $X_1, \dots, X_n$  are conditional independent  $Bern(p)$ , i.e.  $X_i|p \sim Bern(p)$ .

From this we can get that  $\sum_{i=1}^n X_i|p = X$ ,  $X \sim Bin(n, p)$ .

And from the Beta-Binomial conjugacy we can get the posterior distribution of  $p$  given  $X_1 = x_1, \dots, X_n = x_n$  is that:

$$p|X_1 = x_1, \dots, X_n = x_n \sim Beta(1 + \sum_{i=1}^n x_i, 1 + n - \sum_{i=1}^n x_i).$$

And we can see that it only depends on the sum of  $x_i$ .

So with the posterior distribution  $p|X_1 = x_1, \dots, X_n = x_n$ , we can calculate the newly inference:

$$P(X_{n+1} = 1|X_1 + \dots + X_n = k) = \frac{k+1}{n+2}.$$

So above all is the perspective of Beta-Binomial Conjugacy to see the Laplace's law of succession.

## Problem 2

(a) Since  $p \sim \text{Beta}(a, b)$ , so from LOOTS, we can get that

$$E[p^2(1-p)^2] = \int_0^1 \frac{1}{\beta(a, b)} p^{a-1}(1-p)^{b-1} \cdot p^2(1-p)^2 dp = \frac{1}{\beta(a, b)} \int_0^1 p^{a+1}(1-p)^{b+1} dp.$$

And since  $\int_0^1 p^{a+1}(1-p)^{b+1} dp = \beta(a+2, b+2)$ .

$$\text{So } E[p^2(1-p)^2] = \frac{\beta(a+2, b+2)}{\beta(a, b)}.$$

Since  $a, b$  are positive real numbers, so  $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ , similarly,  $\beta(a+2, b+2) = \frac{\Gamma(a+2)\Gamma(b+2)}{\Gamma(a+b+4)}$ .

Also, since  $a, b$  are positive real numbers, so  $\Gamma(a+2) = (a+1)\Gamma(a+1) = (a+1)a\Gamma(a)$ , similarly,  $\Gamma(b+2) = (b+1)\Gamma(b+1) = (b+1)b\Gamma(b)$ .

$$\text{So } \beta(a+2, b+2) = \frac{(a+1)a\Gamma(a)(b+1)b\Gamma(b)}{(a+b+3)(a+b+2)(a+b+1)(a+b)\Gamma(a+b)}.$$

$$\begin{aligned} \text{So } E[p^2(1-p)^2] &= \frac{\beta(a+2, b+2)}{\beta(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{(a+1)a\Gamma(a)(b+1)b\Gamma(b)}{(a+b+3)(a+b+2)(a+b+1)(a+b)\Gamma(a+b)} \\ &= \frac{(a+1)a(b+1)b}{(a+b+3)(a+b+2)(a+b+1)(a+b)}. \end{aligned}$$

$$\text{So above all, } E[p^2(1-p)^2] = \frac{(a+1)a(b+1)b}{(a+b+3)(a+b+2)(a+b+1)(a+b)}.$$

(b) It only depends on the fact that  $A$  won exactly 6 of the 10 games on record.

According to what we have learn about Bayesian Inference, we can get that:

The prior distribution of  $p$  is  $\text{Unif}(0, 1)$ .

And since  $\beta(1, 1) = \frac{\Gamma(1)\Gamma(1)}{\Gamma(2)} = 1$ , so  $\text{Unif}(0, 1) \sim \text{Beta}(1, 1)$ , i.e. the prior distribution of  $p \sim \text{Beta}(1, 1)$ .

And for the historical data, the winners are  $AAABBBABAB$ .

Let  $X_i$  records the  $i$ -th pervious data, if  $A$  won the game, then  $X_i = 1$ , otherwise  $X_i = 0$ .

So  $X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 0, X_6 = 1, X_7 = 1, X_8 = 0, X_9 = 1, X_{10} = 0$ .

And we can get that  $X_i|p \sim \text{Bern}(p)$ , so  $S = \sum_{i=1}^{10} X_i|p \sim \text{Bin}(10, p)$ . And  $S = 6$ .

According to the Beta-Binomial conjugacy, we can get that the posterior distribution of  $p$  only depends on the sum of  $X_i$  i.e.  $S$ .

And we can get that  $S = 6$ , for the total historical data of 10 games,  $A$  won 6 games.

So according to the Beta-Binomial conjugacy, since the prior distribution  $p \sim \text{Beta}(1, 1)$ ,

so we can get that the posterior distribution of  $p$  given  $S = 6$  is that:

$$p|S = 6 \sim \text{Beta}(1+6, 1+(10-6)) \sim \text{Beta}(7, 5).$$

i.e.  $p|\text{historical data} \sim \text{Beta}(7, 5)$ .

So above all, the posterior distribution of  $p$  only depends on the number of wins for  $A$  in the 10 games on record.

(c) From the analysis in (b), we can get that the prior distribution  $p \sim \text{Beta}(0, 1)$ ,

and  $S = \sum_{i=1}^{10} X_i|p \sim \text{Bin}(10, p)$ , so the posterior distribution of  $p$  given  $S = 6$  is that:

$$p|\text{historical data} \sim \text{Beta}(1+6, 1+(10-6)) \sim \text{Beta}(7, 5).$$

So above all, the posterior distribution of  $p$  given the historical data is that  $p|\text{historical data} \sim \text{Beta}(7, 5)$ .

(d) 1. Conditioning on  $p$ , the indicator of  $A$  winning in the first game of the match **uncorrelated with** the indicator of  $A$  winning in the second game of the match.

According to the analysis in (b),(c), and according to what we have learned about Bayesian Inference, with given  $p$ , the first game and the second game are conditionally independent.

So they are uncorrelated.

2. Conditioning on the historical data, the indicator of  $A$  winning in the first game of the match **positively correlated with** the indicator of  $A$  winning in the second game of the match.

Without given  $p$ , we can intuitive regard that winning in the first game of the match could increases the estimation of  $p$ , it can also increase the belief that  $A$  will win the next i.e. second game, and vise versa.

So winning the first game increases the probability of winning the second game that we are estimating, and vise versa.

So they are positively correlated.

(e) Since the match is not yet decided when going into the fifth games, the the first four games must tied, i.e.  $A$  wins 2 games, and  $B$  wins 2 games.

With the given historical data, i.e. with given  $p$ , we can get that  $A$  has the probability of  $p$  to win each game.

So for the situation that  $A$  only wins 2 games in the first four games, the probability that this situation happens is that  $\binom{4}{2}p^2(1-p)^2$ .

And from (c), we have known that  $p \sim \text{Beta}(7,5)$ , and from (a), we get that when  $p \sim \text{Beta}(a,b)$ ,

$$E[p^2(1-p)^2] = \frac{(a+1)a(b+1)b}{(a+b+3)(a+b+2)(a+b+1)(a+b)}.$$

So we can get the expectation of  $\binom{4}{2}p^2(1-p)^2$  is that:

$$\begin{aligned} E[\binom{4}{2}p^2(1-p)^2] &= \binom{4}{2}E[p^2(1-p)^2] \\ &= \binom{4}{2} \frac{(a+1)a(b+1)b}{(a+b+3)(a+b+2)(a+b+1)(a+b)} = \binom{4}{2} \frac{8*7*6*5}{15*14*13*12} = \frac{4}{13} \end{aligned}$$

So above all, given the historical data, the expected value for the probability that the match is not yet decided when going into the fifth games is that  $\frac{4}{13}$ .

### Problem 3

(a) Since  $U_1, \dots, U_n \sim \text{Unif}(0, 1)$ , so  $f(x) = 1, 0 < x < 1$ . And  $F(x) = x, 0 < x < 1$

The joint PDF of  $U_{(1)}, \dots, U_{(n)}$  is  $f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = \lim_{\delta_1, \dots, \delta_n \rightarrow 0} \frac{P(U_{(1)} \in (u_1 - \delta, u_1 + \delta), \dots, U_{(n)} \in (u_n - \delta, u_n + \delta))}{(2\delta_1) \cdots (2\delta_n)}$

For  $U_1, \dots, U_n$ , there are total  $n!$  permutations, and for each permutation, suppose that  $U_i$  be the  $i$ -th order statistic, so we can get that:

$$\lim_{\delta_1, \dots, \delta_n \rightarrow 0} \frac{P(U_1 \in (u_1 - \delta_1, u_1 + \delta_1), \dots, U_n \in (u_n - \delta_n, u_n + \delta_n))}{(2\delta_1) \cdots (2\delta_n)} = f(u_1) \cdots f(u_n) = 1.$$

$$\begin{aligned} \text{So } & \lim_{\delta_1, \dots, \delta_n \rightarrow 0} \frac{P(U_{(1)} \in (u_1 - \delta, u_1 + \delta), \dots, U_{(n)} \in (u_n - \delta, u_n + \delta))}{(2\delta_1) \cdots (2\delta_n)} \\ &= n! \cdot \lim_{\delta_1, \dots, \delta_n \rightarrow 0} \frac{P(U_1 \in (u_1 - \delta_1, u_1 + \delta_1), \dots, U_n \in (u_n - \delta_n, u_n + \delta_n))}{(2\delta_1) \cdots (2\delta_n)} = n!. \end{aligned}$$

i.e.  $f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = n!, u_1 \leq u_2 \leq \dots \leq u_n$ .

And otherwise,  $f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = 0$ .

So above all, the joint PDF of  $U_{(1)}, \dots, U_{(n)}$  is that

$$f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = n!, u_1 \leq u_2 \leq \dots \leq u_n$$

$$f_{U_{(1)}, \dots, U_{(n)}}(u_1, \dots, u_n) = 0, \text{otherwise.}$$

(b) For the joint PDF of  $U_{(j)}$  and  $U_{(k)}$ , since  $1 \leq j < k \leq n$ , considering that they are in a small interval  $U_{(j)} \in (u_j - \delta_1, u_j + \delta_1)$ ,  $U_{(k)} \in (u_k - \delta_2, u_k + \delta_2)$ . And  $u_j < u_k$ .

As for the two variables  $U_{(j)}$  and  $U_{(k)}$ ,

we firstly need choose  $U_{(j)}$  from all  $n$  variables, so the number of ways is  $n$ .

And its range is in  $(u_j - \delta_1, u_j + \delta_1)$ ,  $\delta_1 \rightarrow 0$ ,

$$\text{so } \lim_{\delta_1 \rightarrow 0} \frac{P(U_{(j)} \in (u_j - \delta_1, u_j + \delta_1))}{2\delta_1} = f(u_j) = 1.$$

So the part for  $U_{(j)}$  is  $n$ .

As for the variable  $U_{(k)}$ ,

Then we need to choose one from the rest  $n - 1$  variables, so the number of ways is  $n - 1$ .

And its range is in  $(u_k - \delta_2, u_k + \delta_2)$ ,  $\delta_2 \rightarrow 0$ ,

$$\text{so } \lim_{\delta_2 \rightarrow 0} \frac{P(U_{(k)} \in (u_k - \delta_2, u_k + \delta_2))}{2\delta_2} = f(u_k) = 1.$$

So the part for  $U_{(k)}$  is  $n - 1$ .

And for other order statistics, they are separated into 3 parts:

$U_{(1)}, \dots, U_{(j-1)}$ , their are total  $j - 1$  variables in the range  $(0, u_j)$ ,

$U_{(j+1)}, \dots, U_{(k-1)}$ , their are total  $k - j - 1$  variables in the range  $(u_j, u_k)$ ,

$U_{(k+1)}, \dots, U_{(n)}$ , their are total  $n - k$  variables in the range  $(u_k, 1)$ ,

For the first part, we need to choose  $j - 1$  variables from  $n - 2$  variables(except  $U_{(j)}$  and  $U_{(k)}$ ), so the number of ways is  $\binom{n-2}{j-1}$ .

After that, all chosen ones should be in the range  $(0, u_j)$ , so the probability is  $F(u_j)^{j-1} = u_j^{j-1}$ .

So for the first part is  $\binom{n-2}{j-1} u_j^{j-1}$ .

For the second part, we need to choose  $k - j - 1$  variables from  $n - 2 - (j - 1) = n - j - 1$  variables, so the number of ways is  $\binom{n-j-1}{k-j-1}$ .

After that, all chosen ones should be in the range  $(u_j, u_k)$ ,  
 so the part is  $(F(u_k) - F(u_j))^{k-j-1} = (u_k - u_j)^{k-j-1}$ .  
 So for the second part is  $\binom{n-j-1}{k-j-1}(u_k - u_j)^{k-j-1}$ .

For the third part, we take the remaining  $n - k$  variables,  
 all variables are in the range  $(u_k, 1)$ , so the part is  $(1 - F(u_k))^{n-k} = (1 - u_k)^{n-k}$ .

Combine all the parts, we can get that the joint PDF is that

$$\begin{aligned} & n(n-1)\binom{n-2}{j-1}u_j^{j-1}\binom{n-j-1}{k-j-1}(u_k - u_j)^{k-j-1}(1 - u_k)^{n-k} \\ &= \frac{n(n-1)(n-2)!(n-j-1)!}{(j-1)!(n-2-(j-1))!(k-j-1)!(n-j-1-(k-j-1))!}u_j^{j-1}(u_k - u_j)^{k-j-1}(1 - u_k)^{n-k} \\ &= \frac{n!}{(j-1)!(k-j-1)!(n-k)!}u_j^{j-1}(u_k - u_j)^{k-j-1}(1 - u_k)^{n-k} \end{aligned}$$

And these are in the range  $0 < u_j < u_k < 1$ .

Otherwise, the joint PDF is 0.

So above all, the joint PDF of  $U_{(j)}$  and  $U_{(k)}$  is that

$$\begin{aligned} f_{U_{(j)}, U_{(k)}}(u_j, u_k) &= \frac{n!}{(j-1)!(k-j-1)!(n-k)!}u_j^{j-1}(u_k - u_j)^{k-j-1}(1 - u_k)^{n-k}, 0 < u_j < u_k < 1 \\ f_{U_{(j)}, U_{(k)}}(u_j, u_k) &= 0, \text{ otherwise} \end{aligned}$$

(c) Let  $U_1, \dots, U_n$  be i.i.d.  $Unif(0, 1)$ .

The  $Unif(0, 1)$ 's PDF is  $f(x) = 1$ , and its CDF is  $F(x) = x$ .

And let  $U_{(j)}$  be the  $j$ th order statistic.

1. From what we have learned, we can get that

$$f_{U_{(j)}}(x) = n\binom{n-1}{j-1}f(x)F(x)^{j-1}(1 - F(x))^{n-j} = \frac{n!}{(n-j)!(j-1)!}x^{j-1}(1 - x)^{n-j}$$

i.e.  $U_{(j)} \sim Beta(j, n - j + 1) \sim B$ .

So for  $P(B \leq p)$ , it means that probability that the  $j$ -th order statistic is less than  $p$ .

2. And when  $U_j \leq p$ , we say that the event  $X_j$  is success.

So  $X_j \sim Bern(p)$ .

And  $X_1, \dots, X_n$  are i.i.d., let  $X = X_1 + \dots + X_n$

So  $X \sim Bin(n, p)$ .

So  $P(X \geq j)$  is the probability that the number of success is greater than or equal to  $j$ .

i.e. there are at least  $j$  variables less than  $p$ .

Which is the same as the probability that the  $j$ -th order statistic is less than  $p$ .

So above all,  $P(X \geq j) = P(B \leq p)$  is proved with two different description of same event.

(d) From (c), we get that  $B \sim Beta(j, n - j + 1)$ , i.e.  $f_B(x) = \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1 - x)^{n-j}, 0 < x < 1$ .

$$\text{So } P(B \leq p) = \int_0^p \frac{n!}{(j-1)!(n-j)!}x^{j-1}(1 - x)^{n-j}dx.$$

And we can also get that  $X \sim Bin(n, p)$ .

$$\text{So } P(X \geq j) = \sum_{k=j}^n \binom{n}{k}p^k(1 - p)^{n-k}.$$

We can also get that  $P(B \leq p) = P(X \geq j)$  from (c).

So  $\int_0^p \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j} dx = \sum_{k=j}^n \binom{n}{k} p^k (1-p)^{n-k}$ .

And swap the  $p$  into  $x$ , we can get that

$$\int_0^x \frac{n!}{(j-1)!(n-j)!} t^{j-1} (1-t)^{n-j} dt = \sum_{k=j}^n \binom{n}{k} x^k (1-x)^{n-k}.$$

So above all

$$\int_0^x \frac{n!}{(j-1)!(n-j)!} t^{j-1} (1-t)^{n-j} dt = \sum_{k=j}^n \binom{n}{k} x^k (1-x)^{n-k}$$

for all  $x \in [0, 1]$ , and  $j, n$  are positive integers with  $j \leq n$  have been proved without using calculus.



## Problem 4

Consider the Poisson process.

The Poisson process with rate  $\lambda$  means that if the interval of arriving numbers  $\sim \text{Expo}(\lambda)$ , then the number of arrivals that occur in an interval of length  $t$  is a  $\text{Pois}(\lambda t)$  random variable.

For this problem, we take  $\lambda = 1$ .

Then the interval of arriving time between two person  $i$  and person  $i + 1$  is  $X_i \sim \text{Expo}(1)$ .

And let  $N_t$  be the number of arriving people in the time interval  $[0, t]$ .

And let  $T_n$  be the time that the  $n$ -th person arrived.

Then  $N_t$  and  $T_n$  have a duality:

i.e. if the  $n$ -th person arrived before time  $t$ , then at time  $t$ , there arrived at least  $n$  people.

So we can get that  $\{N_t \geq n\} \Leftrightarrow \{T_n \leq t\}$ .

And since  $T_n$  is the time that the  $n$ -th person arrived, so from the definition, we can get that  $T_n = \sum_{i=1}^n X_i$ .

And from what we have learned, if  $X_1, \dots, X_n$  are i.i.d.  $\text{Expo}(\lambda)$ , then  $X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$ .

And from descriptions above, we can get that  $X_i$  are i.i.d.  $\text{Expo}(1)$ , so  $T_n \sim \text{Gamma}(n, 1)$ .

Also, from the description of Poisson process, we have said that the number of arrivals that occur in an interval of length  $t$  is a  $\text{Pois}(\lambda t)$  random variable.

From this situation, we know that  $\lambda = 1$ , and  $N_t$  is the arrivals in the interval  $[0, t]$ , whose length is  $t$ .

So  $N_t \sim \text{Pois}(t)$ .

Combine the informations above, we have:

1.  $\{N_t \geq n\} \Leftrightarrow \{T_n \leq t\}$
2.  $T_n \sim \text{Gamma}(n, 1)$ .
3.  $N_t \sim \text{Pois}(t)$ .

So we have  $P(N_t \geq n) = P(T_n \leq t)$ .

We can do some replacement to the variables.

Let  $X$  replace  $N_t$ .

Let  $Z$  replace  $T_n$ .

Let  $\lambda$  replace  $t$ .

Let  $k + 1$  replace  $n$ .

With these replacements, we can get that

$X \sim \text{Pois}(\lambda)$  and  $Z \sim \text{Gamma}(k + 1, 1)$ , and

$$\begin{aligned} P(X \geq k + 1) &= P(Z \leq \lambda) \\ 1 - P(X \geq k + 1) &= 1 - P(Z \leq \lambda) \end{aligned}$$

Since  $X$  is a discrete random variable,  $Z$  is a continuous random variable.

So  $1 - P(X \geq k + 1) = P(X \leq k)$ , and  $1 - P(Z \leq \lambda) = P(Z > \lambda)$

i.e.

$$P(X \leq k) = P(Z > \lambda)$$

So above all, when  $X \sim \text{Pois}(\lambda)$  and  $Z \sim \text{Gamma}(k + 1, 1)$ , then

$$P(X \leq k) = P(Z > \lambda)$$

have been proved with Poisson process.

## Problem 5

(a) Let  $X \sim \text{Beta}(2, 4)$ .

We can calculate the beta integral:

$$\beta(2, 4) = \int_0^1 x^{2-1}(1-x)^{4-1}dx = \int_0^1 x(1-x)^3dx = \frac{1}{20}$$

Then we can calculate the PDF of  $X$ :  $f(x) = f_X(x) = \frac{1}{\beta(2, 4)}x^{2-1}(1-x)^{4-1} = 20x(1-x)^3, 0 < x < 1$

As for the acceptance-rejection algorithm.

Since the support of Beta distribution is  $(0, 1)$ , so we can take  $Y \sim \text{Unif}(0, 1)$ .

And its support is  $g(y) = f_Y(y) = 1$ .

Let  $c$  denote a constant s.t.  $c \geq \sup_y \frac{f(y)}{g(y)}$ .

$$\frac{f(y)}{g(y)} = \frac{20y(1-y)^3}{1} = 20y(1-y)^3.$$

Since  $y \in (0, 1)$ , so we can get that when  $y = \frac{1}{4}$ ,  $\frac{f(y)}{g(y)}$  will take the maximum value within the domain.

$$\text{i.e. } \sup_y \frac{f(y)}{g(y)} = \frac{f(\frac{1}{4})}{g(\frac{1}{4})} = \frac{135}{64}.$$

$$\text{i.e. } c \geq \frac{135}{64}.$$

$$\text{So we can just take } c = \frac{135}{64}.$$

Then we can do the acceptance-rejection algorithm.

1. Generate  $Y \sim \text{Unif}(0, 1)$ .

2. Generate  $U \sim \text{Unif}(0, 1)$ .

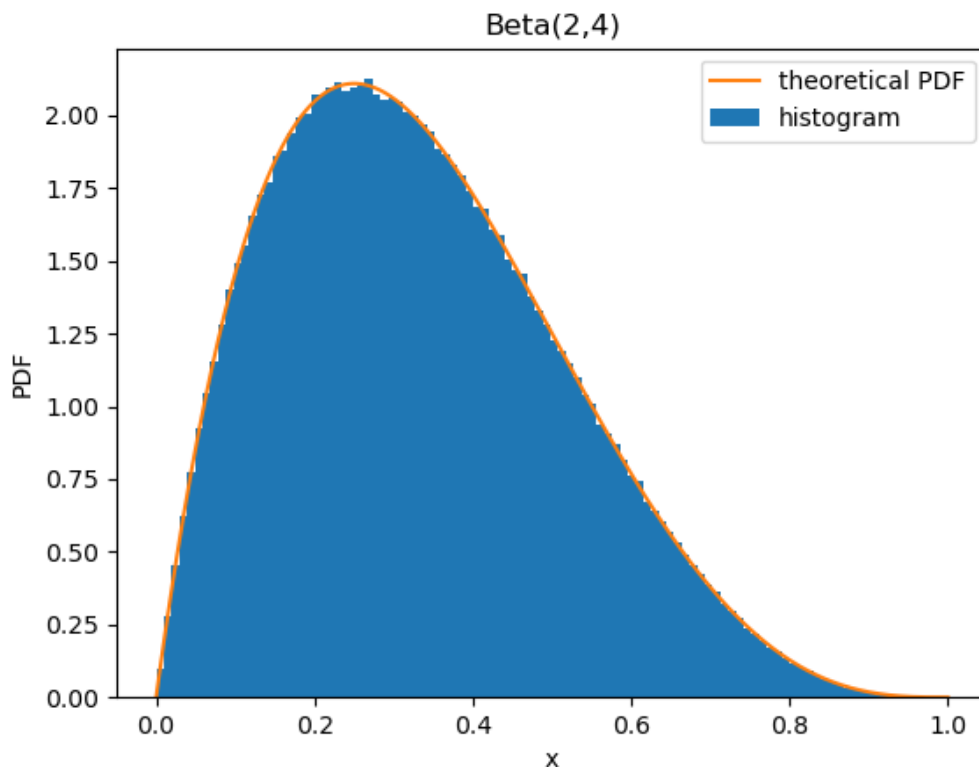
3. If  $U \leq \frac{f(Y)}{c \cdot g(Y)} = \frac{20Y(1-Y)^3}{\frac{135}{64} \cdot 1} = \frac{256}{27}Y(1-Y)^3$ , then set  $X = Y$

4. Else, go back to step 1.

With the above algorithm, we can generate the samples on  $X \sim \text{Beta}(2, 4)$ .

And we can get the following histogram figure, it is compared with the theoretical PDF.

The following is the histogram figure and theoretical PDF of  $\text{Beta}(2, 4)$ :

Figure 1: Histogram and theoretical PDF of  $Beta(2,4)$ 

(b) Similarly with (a), we want to sample on  $N(0, 1)$  with acceptance-rejection algorithm.

Let  $Z \sim N(0, 1)$ , and  $X = |Z|$ .

So  $Z \in (-\infty, +\infty)$ ,  $X \in (0, +\infty)$ .

We can calculate the PDF of  $X$ :

$$f(x) = f_X(x) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x > 0.$$

And we can choose that  $Y \sim Expo(1)$ , and its PDF is  $g(y) = f_Y(y) = e^{-y}, y > 0$ .

Following the method in (a), we can get that  $c \geq \sup_y \frac{f(y)}{g(y)}$ .

$$\frac{f(y)}{g(y)} = \frac{\frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}}{e^{-y}} = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2+y}.$$

Since  $y > 0$ , so we can get that when  $y = 1$ ,  $\frac{f(y)}{g(y)}$  will take the maximum value within the domain.

$$\text{i.e. } \sup_y \frac{f(y)}{g(y)} = \frac{f(1)}{g(1)} = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}+1} = \sqrt{\frac{2e}{\pi}}.$$

So we can just take  $c = \sqrt{\frac{2e}{\pi}}$ .

Then we can do the acceptance-rejection algorithm.

1. Generate  $Y \sim Expo(1)$ .
2. Generate  $U \sim Unif(0, 1)$ .

3. If  $U \leq \frac{f(Y)}{c \cdot g(Y)} = \frac{\frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}Y^2}}{\sqrt{\frac{2e}{\pi}} \cdot e^{-Y}} = \frac{1}{\sqrt{e}} e^{-\frac{1}{2}Y^2 + Y} = e^{-\frac{1}{2}(Y-1)^2}$ , then set  $X = Y$
4. Else, go back to step 1.

After that, we can get the sample the distribution of  $X$ .

To sample on  $Z \sim N(0, 1)$ , since  $X = |Z|$ , so we can generate  $U' \sim \text{Unif}(0, 1)$ ,

$$\text{and let } Z = \begin{cases} X, & U' \leq \frac{1}{2} \\ -X, & U' > \frac{1}{2} \end{cases}.$$

After that, we can sample the distribution of  $Z \sim N(0, 1)$  with acceptance-rejection method.

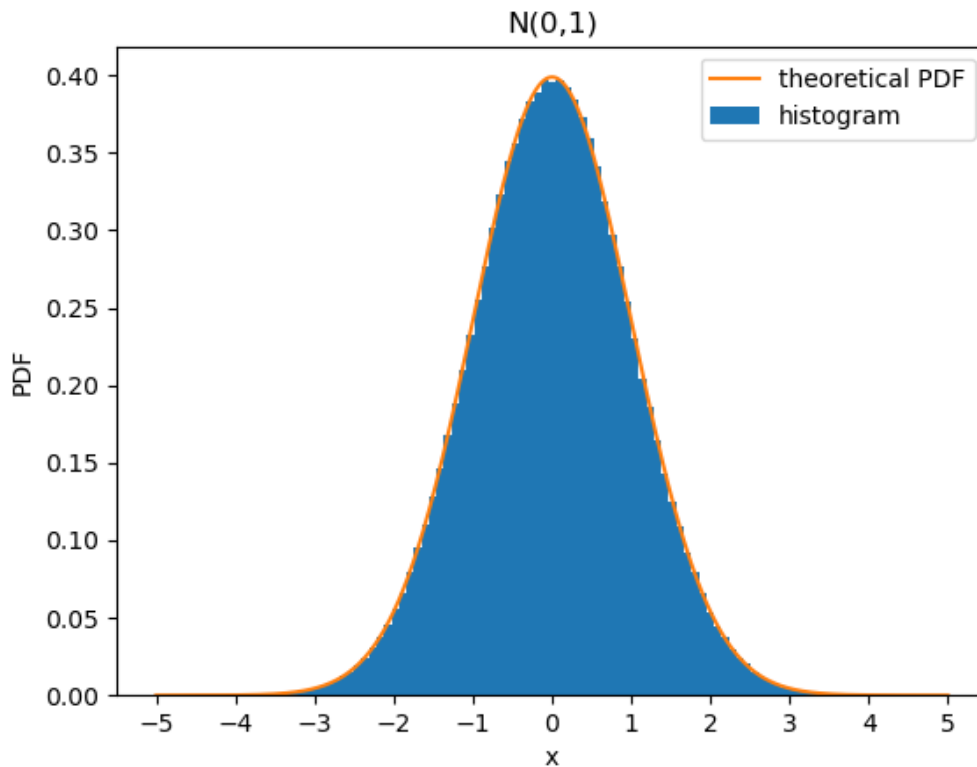


Figure 2: Histogram and theoretical PDF of  $N(0, 1)$

Prove the correctness of the algorithm in theorem:

1. Let event  $A = "U \leq \frac{f(Y)}{c \cdot g(Y)}"$ .

From the decription of the algorithm, we know that we let  $X = Y$  when  $A$  happens.

So the PDF of generated r.v.s.  $X$  is that  $f_Y(y|A) = \frac{P(A|Y=y)}{P(A)} \cdot f_Y(y)$ .

2. And  $P(A|Y=y) = P(U \leq \frac{f(Y)}{c \cdot g(Y)} | Y=y) = P(U \leq \frac{f(y)}{c \cdot g(y)} | Y=y)$ .

Since  $U$  and  $Y$  are independent, so  $P(U \leq \frac{f(y)}{c \cdot g(y)} | Y=y) = P(U \leq \frac{f(y)}{c \cdot g(y)})$ .

And since  $U \sim \text{Unif}(0, 1)$ , so  $P(U \leq \frac{f(y)}{c \cdot g(y)}) = \frac{f(y)}{c \cdot g(y)}$ .

So  $P(A|Y = y) = \frac{f(y)}{c \cdot g(y)}$ .

3. As for  $P(A)$ , with LOTP, we can get that

$$P(A) = \int_0^{+\infty} P(A|Y = y)g(y)dy.$$

From 2., we get that  $P(A|Y = y) = \frac{f(y)}{c \cdot g(y)}$ .

$$\text{So } P(A) = \int_0^{+\infty} \frac{f(y)}{c \cdot g(y)}g(y)dy = \frac{1}{c} \int_0^{+\infty} f(y)dy = \frac{1}{c}.$$

For the last step, this is because  $g(y)$  is the PDF of  $\text{Expo}(1)$ , which support is  $(0, +\infty)$ , so  $\int_0^{+\infty} g(y)dy = 1$ .

4. Combine 2., 3. into 1., we can get that

$$P(Y = y|A) = \frac{P(A|Y = y)}{P(A)} \cdot f_Y(y) = \frac{\frac{f(y)}{c \cdot g(y)}}{\frac{1}{c}} \cdot g(y) = f(y).$$

So from 1. to 4., we have prove that the PDF of generated r.v.s.  $X$  is that  $f_Y(y|A) = f(y)$ .

i.e.  $X \sim f(y)$ .

And since  $X = |Z|$ , so we generated a r.v.s.  $U' \sim (0, 1)$ .

$$\text{And let } Z = \begin{cases} X, & U' \leq \frac{1}{2} \\ -X, & U' > \frac{1}{2} \end{cases}.$$

This is because  $Z \sim (0, 1)$ , so it is symmetric about  $x = 0$ .

And since  $X = |Z|$ , so it can be regard that  $X$  takes the value of  $Z$  with probability  $\frac{1}{2}$ , and takes the value of  $-Z$  with probability  $\frac{1}{2}$ .

So we can generate  $X$  with  $U' \sim (0, 1)$ , and let  $X = Z$  when  $U' \leq \frac{1}{2}$ , and let  $X = -Z$  when  $U' > \frac{1}{2}$ .

So above all, we can generate  $Z \sim N(0, 1)$  with acceptance-rejection method. The correctness have been proved.

(c) Both Acceptance-Rejection method and Box-Muller method can be used to sample on  $N(0, 1)$ . And each of them have their own pros and cons

As for the Acceptance-Rejection method:

- Pros:
  - 1. It is easy to implement.
  - 2. It can be used to sample on any distribution, we do not have to need the exact expression of the distribution that we want to sample, we only need to know the relativeness.
  - 3. It can sample on any distribution, not only  $N(0, 1)$ .
- Cons:
  - 1. It may be difficult to find a suitable constant  $c$ .
  - 2. It may be slower to sample because for each time, the probability of acceptance  $\sim FS(p)$ , where  $p = \frac{1}{c}$ , so the expected sample time is  $\frac{1}{p} = \frac{1}{\frac{1}{c}} = c$ . So it may takes  $c$  times to get one sample compared with Box-Muller method.
  - 3. The quantity of the sample points is not fixed, it depends on the constant  $c$ . If  $c$  is too big, then it may takes a lot of time to get one sample, and if it is too small, it may loss the accuracy.

As for the Box-Muller method:

- Pros:
  - 1. It is easy to implement.
  - 2. It is fast.
  - 3. It is accurate.
  - 4. It is easy to generate two independent  $N(0, 1)$  r.v.s.
- Cons:
  - 1. It can only be used to sample on  $N(0, 1)$ .
  - 2. It requires the exact expression of the distribution that we want to sample.
  - 3. It requires the trigonometric function and exponential function, which is advanced, and may be difficult to implement.

So above all, Acceptance-Rejection method can be used to sample on various distributions, but it may be slow and difficult to find a suitable constant  $c$ .

Box-Muller method can only be used to sample on  $N(0, 1)$ , but it is fast and accurate.

(d) To calculate  $c = P(Y > 8)$ , where  $Y \sim N(0, 1)$ .

So the PDF of  $Y$  is that  $f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ .

It may be difficult to calculate the exact value of  $c$  using simple sampling methods (because of the  $3\sigma$ 's principle, the result must be very small).

So we can use the importance sampling.

Take  $g \sim N(8, 1)$ .

Let  $Y_1, \dots, Y_N \sim g$ , so the PDF of  $g$  is that  $g(y_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_j-8)^2}$ .

Let  $h(Y_j)$  be the indicator that whether  $Y_j > 8$ .

So with monty carlo method, we can get that

$$c = P(Y > 8) = E[I(Y > 8)] = \frac{1}{n} \sum_{j=1}^n I(Y'_j > 8) = \frac{1}{n} \sum_{j=1}^n h(Y'_j), \text{ where } Y'_j \sim N(0, 1).$$

With importance sampling, since  $I = E_f[h(Y)] = \int h(y)f(y)dy = \int \frac{h(y)f(y)}{g(y)}g(y)dy = E_g[\frac{h(Y)f(Y)}{g(Y)}]$ .

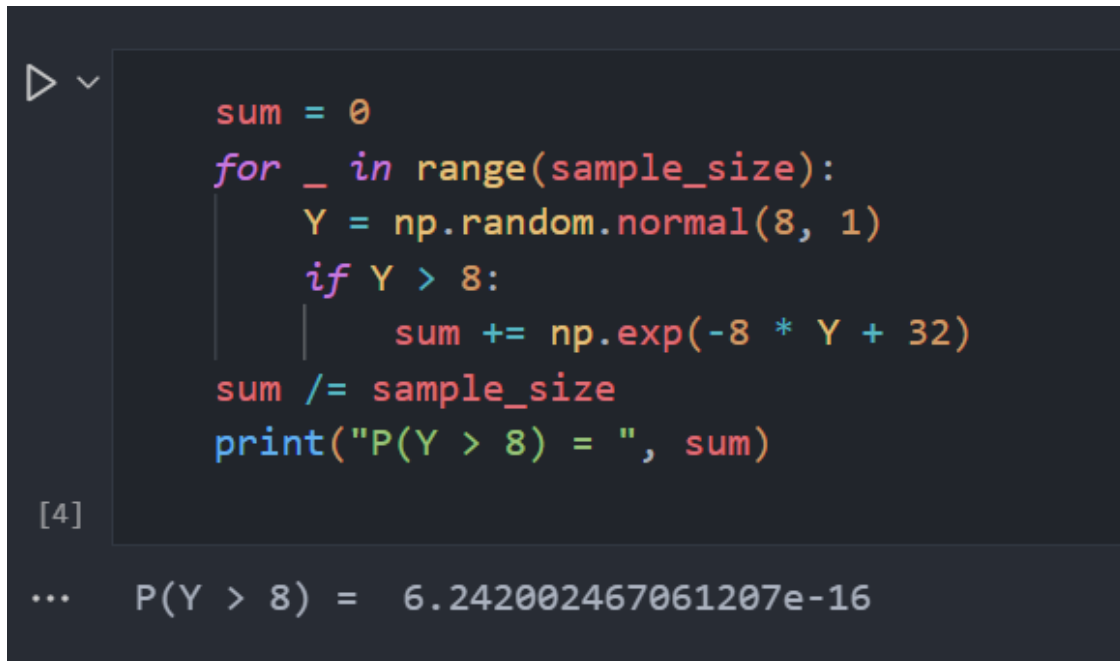
So  $I' = \frac{1}{n} \sum_{i=1}^n \frac{h(Y_j)f(Y_j)}{g(Y_j)}$ , where  $Y_j \sim N(8, 1)$ .

we can get that  $c' = \frac{1}{n} \sum_{j=1}^n \frac{h(Y_j)f(Y_j)}{g(Y_j)}$

$$= \frac{1}{n} \sum_{j=1}^n I(Y_j > 8) \cdot \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Y_j^2}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y_j-8)^2}}$$

$$= \frac{1}{n} \sum_{j=1}^n I(Y_j > 8) \cdot e^{-8Y_j+32}$$

So we just need to sample  $n$  times,  $Y_j \sim N(8, 1)$ , and calculate the average of  $I(Y_j > 8) \cdot e^{-8Y_j+32}$ .



```
sum = 0
for _ in range(sample_size):
    Y = np.random.normal(8, 1)
    if Y > 8:
        sum += np.exp(-8 * Y + 32)
sum /= sample_size
print("P(Y > 8) = ", sum)
```

[4]

... P(Y > 8) = 6.242002467061207e-16

Figure 3:  $P(Y > 8)$ ,  $Y \sim N(0, 1)$

The result using important sampling method is  $6.242002467061207e - 16$ .

Which is very close to the correct answer  $6.25 * 10^{-16}$ .

So we can regard that the importance sampling method is effective and provide correct answers.

The following are the pages from jupyter notebook to show that the code can successful run out the images and calculation results we mentioned above.