Probability & Statistics for EECS: Homework 12 # Solution

Problem 1

Laplace's law of succession says that if $X_1, X_2, \ldots, X_{n+1}$ are conditionally independent Bern(p) r.v.s given p, but p is given a Unif$(0,1)$ prior to reflect ignorance about its value, then

$$P\left(X_{n+1}=1 \mid X_1+\cdots+X_n=k\right)=\frac{k+1}{n+2}$$

As an example, Laplace discussed the problem of predicting whether the sun will rise tomorrow, given that the sun did rise every time for all n days of recorded history; the above formula then gives $(n+1)/(n+2)$ as the probability of the sun rising tomorrow.

(a) Find the posterior distribution of p given $X_1=x_1, X_2=x_2, \ldots, X_n=x_n$, and show that it only depends on the sum of the x_j (so we only need the one-dimensional quantity $x_1+x_2+\cdots+x_n$ to obtain the posterior distribution, rather than needing all n data points).

(b) Prove Laplace's law of succession, using a form of LOTP to find

$$P\left(X_{n+1}=1 \mid X_1+\cdots+X_n=k\right)$$

by conditioning on p.

(c) Reinterpret the Laplace's law of succession from the perspective of Beta-Binomial Conjugacy.

Solution

(a) By the coherency of Bayes' rule (BH Section 2.6: If we receive multiple pieces of information and wish to update our probabilities to incorporate all the information, it does not matter whether we update sequentially, taking each piece of evidence into account one at a time, or simultaneously, using all the evidence at once), we can update sequentially: updated based on $X_1=x_1$, then based on $X_2=x_2$, etc. The prior is Beta$(1,1)$. By conjugacy of the Beta and Binomial, the posterior after observing $X_1=x_1$ is Beta$\left(1+x_1, 1+(1-x_1)\right)$. This becomes the new prior, and then the new posterior after observing $X_2=x_2$ is Beta$\left(1+x_1+x_2, 1+(1-x_1)+(1-x_2)\right)$. Continuing in this way, the posterior after all of $X_1, \ldots, X_n$ have been observed is

$$p \mid X_1=x_1, \ldots, X_n=x_n \sim \operatorname{Beta}\left(1+\sum_{j=1}^n x_j, 1+n-\sum_{j=1}^n x_j\right).$$

This distribution depends only on $\sum_{j=1}^n x_j$, and is the same posterior distribution we would have obtained if we had observed only the r.v. $X_1+\cdots+X_n$.

(b) Let $S_n=X_1+\cdots+X_n$. By LOTP

$$P\left(X_{n+1}=1 \mid S_n=k\right)=\int_0^1 P\left(X_{n+1}=1 \mid p, S_n=k\right) f\left(p \mid S_n=k\right) dp,$$

where $f\left(p \mid S_n=k\right)$ is the posterior PDF of p given $S_n=k$. By (a) and the fact that the X_j are conditionally independent given p, we then have

$$
\begin{aligned}
P\left(X_{n+1}=1 \mid S_n=k\right) &= \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \int_0^1 p p^k (1-p)^{n-k} dp \\
&= \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)} \\
&= \frac{k+1}{n+2},
\end{aligned}
$$

where to evaluate the integral we pattern-matched to the Beta$(k+2, n-k+1)$ PDF.

(c) According to Beta-Binomial Conjugacy, if we know the prior distribution of $p \sim \text{Beta}(a,b)$, $X \mid p \sim \text{Bin}(n,p)$, then the posterior distribution of $p \mid X = k \sim \text{Beta}(a+k, b+n-k)$. In this problem, since $X_1, X_2, \ldots X_n \sim \text{Bern}(p)$ are conditionally independent r.v.s given p, so $X_1+X_2+\ldots X_n \mid p \sim \text{Bin}(n,p)$. So if we know the prior distribution of $p \sim \text{Unif}(0,1)$, which is equal to $p \sim \text{Beta}(1,1)$, we have:

$$p \mid X_1 + X_2 + \ldots + X_n = k \sim \text{Beta}(k+1, n+1-k).$$

And since $X_{n+1} \sim$ Bern(p), so the probability is:

$$P\left(X_{n+1}=1 \mid X_1+\cdots+X_n=k\right) = E(p \mid X_1+\cdots+X_n=k) = \frac{k+1}{n+2}.$$

Problem 2

(a) Let $p \sim \text{Beta}(a, b)$, where a and b are positive real numbers. Find $E(p^2(1-p)^2)$, fully simplified (Γ should not appear in your final answer).

Two teams, A and B, have an upcoming match. They will play five games and the winner will be declared to be the team that wins the majority of games. Given p, the outcomes of games are independent, with probability p of team A winning and $(1-p)$ of team B winning. But you don't know p, so you decide to model it as an r.v., with $p \sim \text{Unif}(0, 1)$ a priori (before you have observed any data).
To learn more about p, you look through the historical records of previous games between these two teams, and find that the previous outcomes were, in chronological order, $AAABBAABAB$. (Assume that the true value of p has not been changing over time and will be the same for the match, though your *beliefs* about p may change over time.)

(b) Does your posterior distribution for p, given the historical record of games between A and B, depend on the specific order of outcomes or only on the fact that A won exactly 6 of the 10 games on record? Explain.

(c) Find the posterior distribution for p, given the historical data.

The posterior distribution for p from (c) becomes your new prior distribution, and the match is about to begin!

(d) Conditional on p, is the indicator of A winning the first game of the match positively correlated with, uncorrelated with, or negatively correlated with the indicator of A winning the second game of the match? What about if we only condition on the historical data?

(e) Given the historical data, what is the expected value for the probability that the match is not yet decided when going into the fifth game (viewing this probability as an r.v. rather than a number, to reflect our uncertainty about it)?

Solution

(a) By LOTUS,

$$\begin{aligned} E\left(p^2(1-p)^2\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^2(1-p)^2 p^{a-1}(1-p)^{b-1} dp \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 p^{a+1}(1-p)^{b+1} dp. \end{aligned}$$

We recognize the integrand as an un-normalized Beta$(a+2, b+2)$ distribution. Multiplying and dividing by a constant and using the fact that PDFs integrate to 1, we obtain

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+2)\Gamma(b+2)}{\Gamma(a+b+4)} = \frac{(a+1)a(b+1)b}{(a+b+3)(a+b+2)(a+b+1)(a+b)},$$

since, e.g., $\Gamma(a+2) = (a+1)\Gamma(a+1) = (a+1)a\Gamma(a)$.

(b) Only the numbers of prior wins and losses for A matter, not the order in which they occurred. By the coherency of Bayes' rule (see Section 2.6), we will get the same posterior distribution from updating all at once on the observed data $AAABBAABAB$ as from updating one game at a time. So consider the one game at a time method. We start with a Beta$(1, 1)$ distribution. Each time A wins, we increment the first parameter by 1; each time B wins, we increment the second parameter by 1. Thus, the posterior distribution depends only on the number of wins for A and the number of wins for B .

(c) The prior on p is Beta$(1,1)$. By conjugacy,

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

(d) These indicator r.v.s are conditionally independent given p , but not independent. Given p , they are therefore uncorrelated. Without being given p, the indicators are positively correlated since learning that A won the first game increases our estimate of p, which in turn increases our degree of belief that A will win the next game.

Here is the mathematical proof. Let I_1 be the indicator of A winning the first game of the match, and I_2 be the indicator of A winning the second game of the match. Conditional on p, I_1 and I_2 are uncorrelated since they are i.i.d. Bernoulli r.v.s when p is given.

If we only condition on the historical data, *i.e.*, $p \sim$ Beta$(7,5)$, we have

$$E\left(I_1\right) = P\left(I_1 = 1\right) = \int_0^1 P\left(I_1 = 1 | p = x\right) f_p\left(x\right) dx = \int_0^1 x f_p\left(x\right) dx = E(p),$$

where the third equation is because $I_1 | p \sim \text{Bern}(p)$. Similarly we also have $E(I_2) = E(p)$. On the other hand,

$$\begin{aligned} E\left(I_1 I_2\right) &= P\left(I_1 I_2 = 1\right) = P\left(I_1 = 1, I_2 = 1\right) \\ &= \int_0^1 \Pr\left(I_1 = 1, I_2 = 1 | p = x\right) f_p\left(x\right) dx = \int_0^1 x^2 f_p\left(x\right) dx = E(p^2) \end{aligned}$$

Since I_1 and I_2 are independent Bernoulli r.v.s with distribution Bern(p) given p. Therefore

$$\text{Cov}\left(I_1, I_2\right) = E\left(I_1 I_2\right) - E\left(I_1\right) E\left(I_2\right) = E(p^2) - E^2(p) = \text{Var}(p) > 0,$$

which implies that I_1 and I_2 are positively correlated.

(e) Given p , the probability that the match is tied after 4 games is $\left(\begin{array}{c} 4 \\ 2 \end{array}\right) p^2(1-p)^2$. But p is unknown, so we will estimate this quantity instead. By (a) and (c), the expected value of $\left(\begin{array}{c} 4 \\ 2 \end{array}\right) p^2(1-p)^2$ given the historical data is

$$\left(\begin{array}{c} 4 \\ 2 \end{array}\right) \cdot \frac{(8)(7)(6)(5)}{(15)(14)(13)(12)} = \frac{4}{13}$$

Problem 3

Let $U_1, \ldots, U_n$ be i.i.d. Unif$(0,1)$. Let $U_{(j)}$ be the corresponding jth order statistic, where $1 \leq j \leq n$.

1. Find the joint PDF of $U_{(1)}, \ldots, U_{(n)}$.
2. Find the joint PDF of $U_{(j)}$ and $U_{(k)}$, where $1 \leq j < k \leq n$.
3. Let $X \sim \text{Bin}(n,p)$ and $B \sim \text{Beta}(j, n-j+1)$, where n is a positive integer and j is a positive integer with $j \leq n$. Show using a story about order statistics that

$$P(X \geq j) = P(B \leq p).$$

   This shows that the CDF of the continuous r.v. B is closely related to the CDF of the discrete r.v. X, and is another connection between the Beta and Binomial.

4. Show that

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

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

Solution

1. Suppose $X_1, \ldots, X_n$ be i.i.d. random variables and their marginal PDF and CDF is $f(x)$ and $F(x)$ respectively. For each $\boldsymbol{x} \in \mathbb{R}^n$ with $x_1 < x_2 < \cdots < x_n$, there are $n!$ permutations of the coordinates of $\boldsymbol{x}$. The PDF of $X_1, X_2, \ldots, X_n$ at each of these points is $f(x_1), f(x_2), \cdots, f(x_n)$ respectively. Hence the joint PDF of $\left(X_{(1)}, X_{(2)}, \ldots, X_{(n)}\right)$ at $\boldsymbol{x}$ is $n!$ times this product:

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1, x_2, \ldots x_n) = n! f(x_1) f(x_2) \cdots f(x_n),$$

   where $x_1 < x_2 < \ldots < x_n$.

   In this problem, $U_1, \ldots, U_n$ be i.i.d. Unif$(0,1)$, we have $F(x) = x$, $f(x) = 1$. Therefore, the joint probability of $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ is:

$$f_{U_{(1)},\ldots,U_{(n)}}(x_1, x_2, \ldots x_n) = n!,$$

   where $x_1 < x_2 < \ldots < x_n$.

2. Suppose $X_1, \ldots, X_n$ be i.i.d. random variables and their marginal PDF and CDF is $f(x)$ and $F(x)$ respectively. We want to compute the probability that $X_{(j)}$ is in an infinitesimal interval dx about x and $X_{(k)}$ is in an infinitesimal interval dy about y. Note that there must be $j-1$ sample variables that are less than x, one variable in the infinitesimal interval about x, $k-j-1$ sample variables that are between x and y, one variable in the infinitesimal interval about y, and $n-k$ sample variables that are greater than y. The number of ways to select the variables is the multinomial coefficient

$$\binom{n}{j-1, 1, k-j-1, 1, n-k} = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}$$

   By independence, the probability that the chosen variables are in the specified intervals is

$$[F(x)]^{j-1} f(x) dx [F(y) - F(x)]^{k-j-1} f(y) dy [1 - F(y)]^{n-k}$$

Therefore, the joint probability of $X_{(j)}$ in an infinitesimal interval dx about x and $U_{(k)}$ in an infinitesimal interval dy about y is:

$$f_{X_{(j)},X_{(k)}}(x,y)dxdy = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}[F(x)]^{j-1}f(x)dx[F(y)-F(x)]^{k-j-1}f(y)dy[1-F(y)]^{n-k}.$$

So the joint PDF of $X_{(j)}, X_{(k)}$ is:

$$f_{X_{(j)},X_{(k)}}(x,y) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}[F(x)]^{j-1}f(x)[F(y)-F(x)]^{k-j-1}f(y)[1-F(y)]^{n-k},$$

where $y > x$.

In this problem, $U_1, \ldots, U_n$ be i.i.d. $\text{Unif}(0,1)$, we have $F(x) = x$, $f(x) = 1$. Therefore, the joint probability of $U_{(j)}$ and $U_{(k)}$ is:

$$f_{U_{(j)},U_{(k)}}(x,y) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!}x^{j-1}(y-x)^{k-j-1}(1-y)^{n-k}.$$

where $y > x$.

3. Let $U_1, \ldots, U_n$ be i.i.d. $Unif(0,1)$. Think of these as Bernoulli trials, where U_j is defined to be "successful" if $U_j \leq p$ (so the probability of success is p for each trial). Let X be the number of successes. Then $X \geq j$ is the same event as $U_{(j)} \leq p$, so $P(X \geq j) = P\left(U_{(j)} \leq p\right)$. By BH Example 8.6.5, $U_{(j)} \sim \text{Beta}(j, n-j+1)$, which is the same distribution as B. Therefore, the event of $X \geq j$ is the same event as $B \leq p$, so we have:

$$P(X \geq j) = P(B \leq p).$$

4. Let $U_1, \ldots, U_n$ be i.i.d. $\text{Unif}(0,1)$ r.v.s, and fix $x \in [0,1]$. Define "success" to mean being at most x and "failure" to mean being greater than x. The lefthand side of the identity is the probability of having at least j successes (since it is the sum of the $\text{Bin}(n,x)$ PMF from j to n), which is:

$$\sum_{k=j}^{n} \binom{n}{k} x^k (1-x)^{n-k},$$

The righthand side is $P\left(U_{(j)} \leq x\right)$, which is

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

since $U_{(j)} \sim \text{Beta}(j, n-j+1)$. Because having at least j successes is the same thing as having $U_{(j)} \leq x$, so the two sides are equal.

Problem 4

If $X \sim \text{Pois}(\lambda)$, $Z \sim \text{Gamma}(k+1, 1)$, where k is a nonnegative integer. Show the Poisson-Gamma Duality holds:

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

Hint: Two possible methods, where one is based on the identity in Problem 3.4, the other is based on the model of Poisson process.

Solution

1. Recursion: Since $X \sim \text{Pois}(\lambda)$, $Z \sim \text{Gamma}(k+1, 1)$, the identity $P(X \leq k) = P(Z > \lambda)$ is equal to:

$$\sum_{j=0}^{k} \frac{\lambda^k}{j!} e^{-x} = \int_{\lambda}^{\infty} \frac{x^k}{k!} e^{-x} dx$$

   Let $f(k, \lambda) = \int_{\lambda}^{\infty} \frac{x^k}{k!} e^{-x} dx$ and we can have the recursion below:

$$\begin{aligned} f(k, \lambda) &= \int_{\lambda}^{\infty} \frac{x^k}{k!} e^{-x} dx \\ &= -\int_{\lambda}^{\infty} \frac{x^k}{k!} de^{-x} \\ &= -(\frac{x^k}{k!} e^{-x} \Big|_{\lambda}^{\infty} - \int_{\lambda}^{\infty} e^{-x} d\frac{x^k}{k!}) \\ &= \frac{\lambda^k}{k!} e^{-\lambda} + \int_{\lambda}^{\infty} \frac{x^{k-1}}{(k-1)!} e^{-x} dx \\ &= \frac{\lambda^k}{k!} e^{-\lambda} + f(k-1, \lambda) \end{aligned}$$

   And since $f(0, \lambda) = e^{-\lambda}$, so

$$f(k, \lambda) = \sum_{j=0}^{k} \frac{\lambda^k}{j!} e^{-x}$$

   Therefore, we prove the identity.

2. Possion Process: Suppose $\{N_t, t \geq 0\}$ is a Poisson process with intensity parameter λ, $N_t \sim \text{Pois}(\lambda t)$. Then the inter-arrival times $X_1, X_2, \cdots \sim \text{Expo}(\lambda)$ (X_i is the inter-arrival time between the arrival of N_{i-1} and N_i, $i = 1, \ldots, n$.) are i.i.d. exponential random variables with mean $1/\lambda$. If W_n denotes the waiting time for the n^{th} event to occur, then

$$T_n = X_1 + X_2 + \cdots + X_n$$

   and $T_n \sim Gamma(n, \lambda)$. Then, we have, $T_n > t \Leftrightarrow N_t < n$.

$$P(T_n > t) = P(N_t < n) = \sum_{x=0}^{n-1} \frac{(\lambda t)^x}{x!} e^{\lambda t}$$

   Let $t = 1$, we can derive the following identity:

$$P(X < k+1) = P(X \leq k) = P(Z/\lambda > 1) = P(Z > \lambda)$$

   If $X \sim Pois(\lambda)$ and $Z/\lambda \sim Gamma(k+1, \lambda)$ (which is equal to $Z \sim Gamma(k+1, 1)$).

3. Based on the identity in 3(d): Let the r.v. $B \sim \text{Beta}(k+1, n-(k+1)+1), Y \sim \text{Bin}(n,p)$ and we have the following due to the above identity in (c),

$$P(B \leq p) = P(Y \geq k+1) \implies P(B \geq p) = P(Y \leq k)$$

Therefore, we have

$$P(Y \leq k) = \int_p^1 \frac{n!}{k!(n-k-1)!} t^k (1-t)^{n-k-1} dt$$

Let $t = x/n$,

$$\begin{aligned}
P(Y \leq k) =& \frac{n!}{k!(n-k-1)!} \times \int_{np}^{n} \left(\frac{x}{n}\right)^k \left(1 - \frac{x}{n}\right)^{n-k-1} \frac{1}{n} dx \\
=& \frac{(n-1)!}{k!(n-1-k)!} \times \int_{np}^{n} \left(\frac{x}{n}\right)^k \left(1 - \frac{x}{n}\right)^{n-k-1} dx \\
=& \int_{np}^{n} \binom{n-1}{k} \left(\frac{x}{n}\right)^k \left(1 - \frac{x}{n}\right)^{n-1-k} dx \\
=& \int_{np}^{n} P(A = k) dx,
\end{aligned}$$

where $A \sim \text{Bin}(n-1, x/n)$ and $P(A = k)$ is its PMF (with value at point k).

Now we consider the "Poisson approximation to Binomial". That is, as $n \to \infty, p \to 0$ and fix $\lambda = np$, we have the following approximations:

- Approximate $Y \sim \text{Bin}(n,p)$ with $X \sim \text{Pois}(\lambda)$;
- Approximate $A \sim \text{Bin}(n-1, \frac{x}{n})$ with $C \sim \text{Pois}(x)$ (since $\frac{x}{n} \to 0$ and $n - 1 \approx n$ when $n \to \infty$)

Therefore, we have

$$\begin{aligned}
&\text{LHS} = P(Y \leq k) = P(X \leq k); \\
&\text{RHS} = \int_\lambda^\infty P(A = k) dx = \int_\lambda^\infty P(C = k) dx = \int_\lambda^\infty \frac{e^{-x} x^k}{k!} dx.
\end{aligned}$$

Since $Z \sim \text{Gamma}(k+1, 1)$, its PDF satisfies

$$f_Z(x) = \frac{1}{\Gamma(k+1)} \times (1 \cdot x)^{k+1} e^{-1 \cdot x} \times \frac{1}{x} = \frac{e^{-x} x^k}{k!}.$$

Therefore, we have

$$P(X \leq k) = \int_\lambda^\infty \frac{e^{-x} x^k}{k!} dx = \int_\lambda^\infty f_Z(x) dx = P(Z > \lambda).$$

Problem 5

Programming Assignment:

(a) Use the Acceptance-Rejection Method to obtain the samples from distribution Beta$(2, 4)$. You need to plot the pictures of both histogram and the theoretical PDF.

(b) Use the Acceptance-Rejection Method to obtain the samples from the standard Normal distribution $\mathcal{N}(0, 1)$. You are required to show the correctness of your algorithm in theory.

(c) Both the Acceptance-Rejection Method and Box-Mulller Method can obtain the samples from the standard Normal distribution $\mathcal{N}(0, 1)$. Discuss the pros and cons of such two methods.

(d) Use the importance sampling method to evaluate the probability of rare event $c = P(Y > 8)$, where $Y \sim N(0, 1)$.

Solution

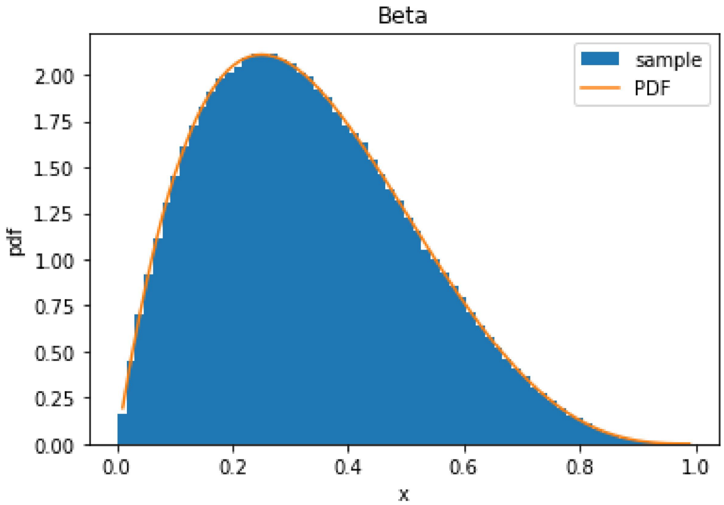
Problem 5 continued on next page...

In [ ]:
```
import random
import numpy as np
import math
import matplotlib.pyplot as plt
```

In [ ]:
```
def plot_data(data):
    plt.hist(data,bins=64,density=True,label="sample")
    plt.xlabel("x")
    plt.ylabel("pdf")
```

(a) Sampling from the Beta Distribution by Acceptance-Rejection Method

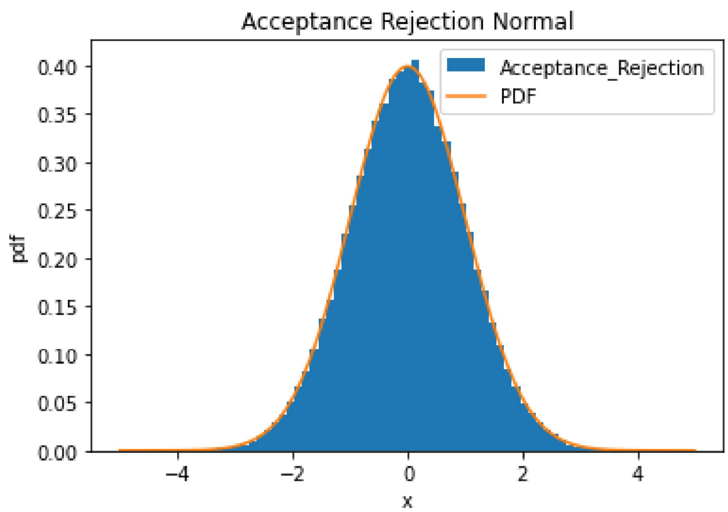
In [ ]:
```
def beta(a,b,n):
    i = 0
    result = []
    while(i < n):
        x = random.uniform(0,1)
        u = random.uniform(0,1)

        # acceptance-rejection
        beta_value = math.factorial(a-1)*math.factorial(b-1) \
            / math.factorial(a+b-1)
        c = (1/beta_value)*pow(((a-1)/(a+b-2)),a-1) \
            *pow(((b-1)/(a+b-2)),b-1)

        if u <= pow(x,a-1)*pow(1-x,b-1)/(pow(((a-1)/(a+b-2)),a-1) \
                                            *pow(((b-1)/(a+b-2)),b-1)):
            i += 1
            result.append(x)

    return result

# parameter (Acceptance and rejection method cannot solve
# when a + b = 2 because the denominator is 0)
a = 2
b = 4
n = 1000000
plot_data(beta(a,b,n))

# corresponding pdf
from scipy.stats import beta
x = np.arange(0.01, 1, 0.01)
y = beta.pdf(x, a, b)
plt.plot(x, y, label='PDF')
plt.title('Beta')
plt.xlabel('x')
plt.ylabel('pdf')
plt.legend()
plt.show()
```

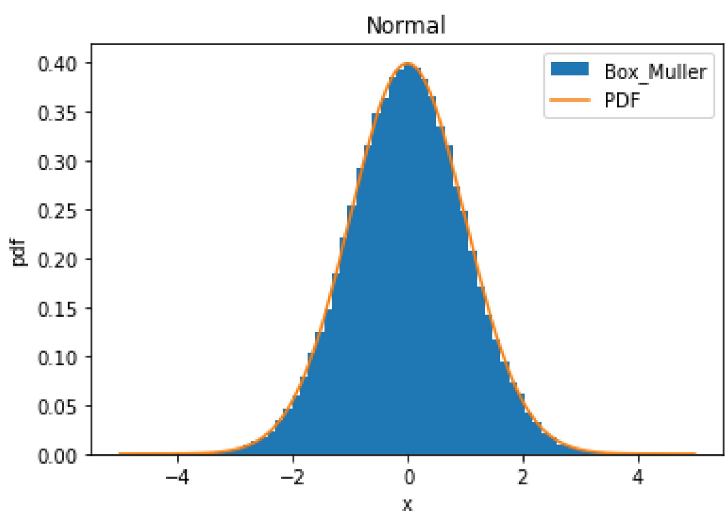


(b) Sampling from the Beta Distribution by Acceptance-Rejection Method and Box-Muller Method

In [ ]:
```
#box muller
def box_muller_normal(n):
    result = []
    for i in range(n):
        x = random.uniform(0,1)
        y = random.uniform(0,1)

        U = math.cos(2*math.pi*x)*math.sqrt(-2*math.log(y))
        V = math.sin(2*math.pi*x)*math.sqrt(-2*math.log(y))
        result.append(U)
    return result

#acceptance and rejection
def acc_rej_norm(n):
    i = 0
    result = []
    while i < n:
        x = np.random.exponential(1)
        u = np.random.uniform(0,1)

        if u <= math.exp((-1/2)*((x-1)**2)):
            i += 1
            z = random.uniform(0,1)
            if z < 0.5:
                result.append(x)
            else:
                result.append(-x)

    return result

n = 100000

plt.hist(acc_rej_norm(n),bins=64,density=True,label="Acceptance_Rejection")
# corresponding pdf
from scipy.stats import norm
x = np.arange(-5, 5, 0.01)
y = norm.pdf(x, 0, 1)
plt.plot(x,y,label="PDF")
plt.title('Acceptance Rejection Normal')
plt.xlabel('x')
plt.ylabel('pdf')
plt.legend()
```

```
plt.show()

plt.hist(box_muller_normal(n),bins=64,density=True,label="Box_Muller")

# corresponding pdf
from scipy.stats import norm
x = np.arange(-5, 5, 0.01)
y = norm.pdf(x, 0, 1)
plt.plot(x,y,label="PDF")
plt.title('Normal')
plt.xlabel('x')
plt.ylabel('pdf')
plt.legend()
plt.show()
```





(c) Pros and cons of Box-Muller and Acceptance-Rejection Methods

Box-Muller:

1. pros: It is easy to implement, and the method only uses Unif(0,1) as the basis data sample, which is simple to sample.

2. cons: Only the standard normal distribution can be sampled by this method.

Acceptance-Rejection:

1. pros: It can sample many kinds of probability distribution including many distributions that is difficult to sample directly.

2. cons: The domain of function g(x) must cover the domain of function f(x). If c is closed to 1, the basis distribution g is still difficult to sample; while if c is closed to 0, the probability of acceptance success will be small, which will cause low effiency.

(d) Importance Sampling to Evaluate the Probability

In [ ]:

```
def importance_sampling(n):
    sample = np.random.normal(8,1,n)
    sum = 0

    for i in range(len(sample)):
        if(sample[i] > 8):
            sum += math.exp(8*(4-sample[i]))

    return sum/n

def without_importance_sampling(n):
    sample = np.random.normal(0,1,n)
    sum = 0

    for i in range(len(sample)):
        if sample[i] > 8:
            sum += 1

    return sum/n

n = 10000000
print("with importance sampling: {}".format(importance_sampling(n)))
print("without sampling sampling: {}".format(without_importance_sampling(n)))
```

```
with importance sampling: 6.227968200258029e-16
without sampling sampling: 0.0
```