STAT2372 Assignment

Ze Hong Zhou

2022-05-16 (last edited 2022-05-19)

```
library(tidyverse)
```

The original source code for this assignment can be found here after I make the repo public.

Question 1

(a) **Let X be a random variable with mean 0 and finite variance σ^2. By applying Markov's inequality to the random variable $W = (X + t)^2$, where $t > 0$, show that**

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2} \quad \text{for any } a > 0.$$

$$
\begin{aligned}
P(X \geq a) &= P(X + t > a + t) \\
&\leq P((X + t)^2 > (a + t)^2) \\
&\leq \frac{E((X + t)^2)}{(a + t)^2} \quad \text{from applying Markov's inequality} \\
&= \frac{E(X^2) + 2tE(X) + t^2}{(a + t)^2} \\
&= \frac{\sigma^2 + t^2}{(a + t)^2} \quad \text{since } E(X) = 0.
\end{aligned}
$$

Let $f(t) = \frac{\sigma^2 + t^2}{(a+t)^2}$. Let the stationary point of $f(t)$ occur when $t = t_m$.

$$
\begin{aligned}
0 &= f'(t_m) \\
&= \frac{(a + t_m)^2 \cdot 2t_m - (\sigma^2 + t_m^2)(2(a + t_m))}{(a + t_m)^4} \\
0 &= (a + t_m)^2 \cdot 2t_m - (\sigma^2 + t_m^2)(2(a + t_m)) \\
&= (a + t_m)(2at_m + 2t_m^2 - 2\sigma^2 - 2t_m^2) \\
&= 2(a + t_m)(at_m - \sigma^2) \\
t_m &= \frac{\sigma^2}{a} \quad \text{since } t > 0.
\end{aligned}
$$

Substituting $t = t_m$ into the inequality,

$$P(X \geq a) \leq \frac{\sigma^2 + \frac{\sigma^4}{a^2}}{(a + \frac{\sigma^2}{a})^2}$$

$$= \frac{\frac{\sigma^2}{a}(a + \frac{\sigma^2}{a})}{(a + \frac{\sigma^2}{a})^2}$$

$$= \frac{\frac{\sigma^2}{a}}{a + \frac{\sigma^2}{a}}$$

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

(b) Hence show that, for any $a > 0$,

$$P(Y \geq \mu + a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P(Y \leq \mu - a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

where $E(Y) = \mu$, $Var(Y) = \sigma^2$.

$$E(Y - \mu) = E(Y) - \mu = 0.$$
$$Var(Y - \mu) = Var(Y) = \sigma^2.$$

Similarly,

$$E(\mu - Y) = 0.$$
$$Var(\mu - Y) = \sigma^2.$$

Applying the result in part (a),

$$P(Y \geq \mu + a) = P(Y - \mu \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

$$P(Y \leq \mu - a) = P(\mu - Y \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

Question 2

(a) Let X have a Poisson distribution with parameter λ.

    (i) Determine $K_X(t)$, the cumulant generating function. Hence find the third and the fourth central moments of X.

$$K_X(t) = \log(E(e^{tX}))$$

$$= \log\left(\sum_{k=0}^{\infty}\left(e^{tk}\frac{e^{-\lambda}\lambda^k}{k!}\right)\right)$$

$$= \log\left(e^{-\lambda}\sum_{k=0}^{\infty}\left(\frac{(\lambda e^t)^k}{k!}\right)\right)$$

$$= \log\left(e^{-\lambda}e^{\lambda e^t}\right) \quad \text{(series expansion of an exponential function)}$$

$$= \lambda(e^t - 1).$$

$$\mu_3 = K_X^{(3)}(0) \qquad\qquad\qquad \sigma^2 = K_X^{(2)}(0)$$
$$= \lambda e^0 \qquad\qquad\qquad\qquad = \lambda$$
$$\mu_3 = \lambda. \qquad\qquad\qquad \mu_4 - 3\sigma^4 = K_X^{(4)}(0)$$
$$\qquad\qquad\qquad\qquad\qquad \mu_4 = \lambda + 3\lambda^2.$$

(ii) Show that the moment-generating function of $Y = (X - \lambda)/\sqrt{\lambda}$ is given by

$$M_Y(t) = \exp\left(\lambda e^{t/\sqrt{\lambda}} - \sqrt{\lambda}t - \lambda\right).$$

$$M_Y(t) = E\left(\exp\left(\frac{X-\lambda}{\sqrt{\lambda}}t\right)\right)$$

$$= E\left(\exp\left(\frac{Xt}{\sqrt{\lambda}}\right)\cdot\exp\left(-\sqrt{\lambda}t\right)\right)$$

$$= e^{-\sqrt{\lambda}t}\sum_{k=0}^{\infty}\left(e^{\frac{kt}{\sqrt{\lambda}}}\frac{e^{-\lambda}\lambda^k}{k!}\right)$$

$$= e^{-\sqrt{\lambda}t-\lambda}\sum_{k=0}^{\infty}\left(\frac{\left(e^{\frac{t}{\sqrt{\lambda}}}\lambda\right)^k}{k!}\right)$$

$$= e^{-\sqrt{\lambda}t-\lambda}\exp\left(e^{\frac{t}{\sqrt{\lambda}}}\lambda\right)$$

$$= \exp\left(\lambda e^{t/\sqrt{\lambda}} - \sqrt{\lambda}t - \lambda\right).$$

(iii) Use the expansion

$$e^{t/\sqrt{\lambda}} = \sum_{i=0}^{\infty}\left(\frac{(t/\sqrt{\lambda})^i}{i!}\right)$$

to show that

$$\lim_{\lambda\to\infty}\left(M_Y(t)\right) = e^{t^2/2}$$

and hence show that the distribution function of Y converges to a standard normal distribution function as $\lambda \to \infty$.

3

$$\lim_{\lambda \to \infty} (M_Y(t)) = \lim_{\lambda \to \infty} \left(\exp\left[\lambda e^{t/\sqrt{\lambda}} - \sqrt{\lambda}t - \lambda \right] \right)$$

$$= \lim_{\lambda \to \infty} \left(\exp\left[\lambda \sum_{i=0}^{\infty} \left(\frac{(t/\sqrt{\lambda})^i}{i!} \right) - \sqrt{\lambda}t - \lambda \right] \right)$$

$$= \lim_{\lambda \to \infty} \left(\exp\left[\lambda \sum_{i=2}^{\infty} \left(\frac{(t/\sqrt{\lambda})^i}{i!} \right) \right] \right)$$

$$= \lim_{\lambda \to \infty} \left(\exp\left[t^2 \sum_{i=2}^{\infty} \left(\frac{(t/\sqrt{\lambda})^{i-2}}{i!} \right) \right] \right)$$

$$= \exp\left[t^2 \lim_{\lambda \to \infty} \left(\sum_{i=2}^{\infty} \left(\frac{(t/\sqrt{\lambda})^{i-2}}{i!} \right) \right) \right]$$

$$= \exp\left[t^2 \left(1/2 + \lim_{\lambda \to \infty} \left[\sum_{i=3}^{\infty} \left(\frac{(t/\sqrt{\lambda})^{i-2}}{i!} \right) \right] \right) \right]$$

$$= e^{t^2/2}.$$

This is the mgf of a standard normal distribution. That is, $\lim_{\lambda \to \infty}(M_Y(t)) = M_Z(t)$. The random variable Y converges in distribution to a standard normal random variable as $\lambda \to \infty$.

(b) X, the number of accidents per year at a given intersection, is assumed to have a Poisson distribution. Over the past few years, an average of 36 accidents per year have occurred at this intersection. If the number of accidents per year is at least 45, an intersection can qualify to be redesigned under an emergency program set up by the state. Approximate the probability that this intersection will come under the emergency program at the end of the next year.

Let $X \sim \text{Pois}(36)$.

$$P(X \geq 45) = 1 - P(X \leq 44)$$
$$= 0.0819.$$

Hence, the probability that this intersection will come under the emergency program at the end of the next year is 0.0819.

Inline code used:

```
(1 - ppois(44, 36)) %>% round(4) %>% format(nsmall = 4)
```

```
## [1] "0.0819"
```

(c) Suppose V_i, $i \in [1, n] \cap \mathbb{N}$, are independent exponential random variables with rate 1. Denote

$$X = \max\left(\left\{ n : \sum_{i=1}^{n} (V_i) \leq \lambda \right\} \right),$$

so X can be thought of as being the maximum number of exponentials having rate 1 that can be summed and still be less than or equal to λ.

(i) Using properties of a Poisson process with rate 1, explain why X has a Poisson distibution with parameter λ.

The exponential distribution with parameter 1 can be defined as the distribution for the random variable that represents the time between two successive events in the Poisson process with parameter 1. Let $\{Y(t) : t \geq 0\}$ be this particular Poisson process. Thus, the random variable X represents the number of times that the time between two successive events has been observed within a time period of λ in $\{Y(t) : t \geq 0\}$. Since this is the same as the number of events that have occurred within this time period, $X \sim Y(t_0 + \lambda) - Y(t_0) \sim \text{Pois}(\lambda)$, where $t_0 > 0$.

(ii) Let $V_i = -log(U_i)$, $U_i \sim$ Uniform$(0,1)$, for all $i \in [1, n] \cap \mathbb{N}$. Show that

$$X = \max\left(\left\{n : \prod_{i=1}^{n}(U_i) \geq e^{-\lambda}\right\}\right), \tag{1}$$

where $\prod_{i=1}^{0}(U_i) = 1$.

$$\lambda \geq \sum_{i=1}^{n}(V_i)$$

$$= \sum_{i=1}^{n}(-\log(U_i))$$

$$= -\log\left(\prod_{i=1}^{n}(U_i)\right)$$

$$\log\left(\prod_{i=1}^{n}(U_i)\right) \geq -\lambda$$

$$\prod_{i=1}^{n}(U_i) \geq e^{-\lambda}.$$

Hence,

$$X = \max\left(\left\{n : \sum_{i=1}^{n}(V_i) \leq \lambda\right\}\right)$$

$$= \max\left(\left\{n : \prod_{i=1}^{n}(U_i) \geq e^{-\lambda}\right\}\right).$$

(iii) It can be shown that (1) is equivalent to

$$X = \min\left(\left\{n : \prod_{i=1}^{n}(U_i) < e^{-\lambda}\right\}\right) - 1.$$

This result may be used to simulate a Poisson random variable with parameter λ. If we continue generating Uniform(0, 1) random variables U_i until their product falls below $e^{-\lambda}$, then the number required, minus 1, is Poisson with parameter λ. Implement this procedure in R to generate 1000 realizations of a Poisson random variable with parameter $\lambda = 5$. Produce the histogram of the generated values $\{x_i : i \in [1, 1000] \cap \mathbb{N}\}$. Find the mean of these values, and compare with its theoretical counterpart.

5

```
set.seed(46375058)
elam <- exp(-5) # e to the power of negative lambda
x_i <- rep(1 - 1, times = 1000) # all x_i start at 0
prod <- runif(1000, 0, 1) # starting test values
logic_test <- prod >= elam # check which x_i needs adjustment

while (FALSE == all(!logic_test)) {
  t <- sum(logic_test) # number of x_i that will increase by 1
  x_i[logic_test] <- x_i[logic_test] + rep(1, times = t) # increase required x_i by 1
  prod[logic_test] <- prod[logic_test] * runif(t, 0, 1) # change required test values
  logic_test <- prod >= elam # check which x_i needs adjustment
}

ggplot(tibble(x_i = x_i)) +
  geom_histogram(aes(x = x_i, y = ..count../sum(..count..)), binwidth = 1) +
  labs(title = "Histogram of Simulated Values of X",
       x = "value", y = "proportion")
```


Histogram of Simulated Values of X
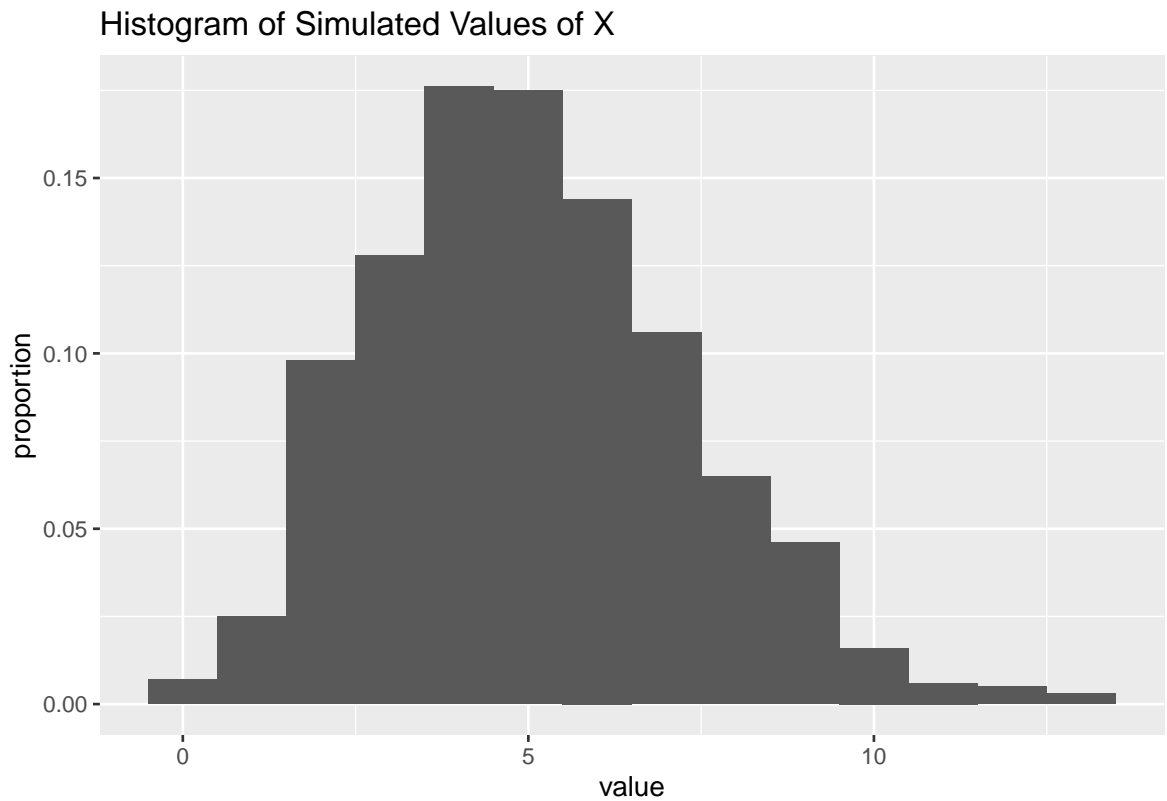
```
(mu <- mean(x_i))
```

```
## [1] 5.049
```

The mean of the simulation data is 5.0490, while the theoretical mean is 5. The relative error is 0.0098. A low absolute relative error and a histogram shaped like a Poisson distribution provides empirical evidence that X is distributed as a Poisson distribution.

Question 3