Final Assessment PHY 125.3

Prasham Bhuta

Final Assessment - Probability

This is the final assessment to the probability chapter of Data Science course at Harvardx.

Loading necessary libraries.

```
library(tidyverse)
library(dslabs)
```

Loading the '2015 US Period Life Table (`death_prob`)' from dslabs and summarising it.

```
data("death_prob")
head(death_prob)
```

```
##   age  sex     prob
## 1   0 Male 0.006383
## 2   1 Male 0.000453
## 3   2 Male 0.000282
## 4   3 Male 0.000230
## 5   4 Male 0.000169
## 6   5 Male 0.000155
```

```
summary(death_prob)
```

```
##       age              sex           prob
##  Min.   :  0.00   Female:120   Min.   :0.000091
##  1st Qu.: 29.75   Male  :120   1st Qu.:0.001318
##  Median : 59.50                Median :0.008412
##  Mean   : 59.50                Mean   :0.127254
##  3rd Qu.: 89.25                3rd Qu.:0.138332
##  Max.   :119.00                Max.   :0.899639
```

The scenario

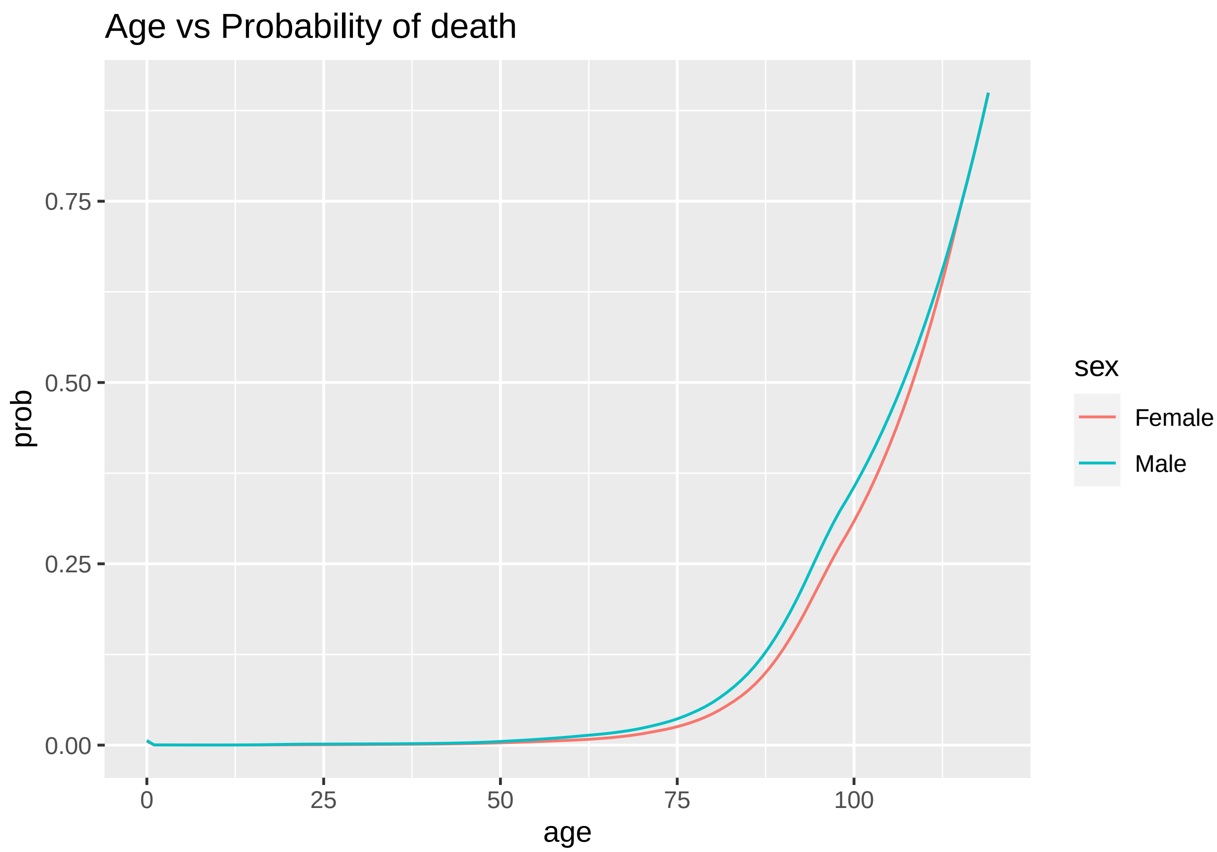
Insurance company offers a 1 year term policy that pays $150000 in event of death. The premium for a 50 year old female is $1,150 yearly. In event of a claim, the company forfeits the premium, and pays $150000 as agreed. The company plans to sell 1000 such insurances.

Lets setup the basic parameters.

```
n <- 1000
loss_payout <- -150000
profit_premium <- 1150
```

Lets plot age vs probability of death.

```
death_prob %>% ggplot(aes(age, prob, group = sex, color = sex)) +
    geom_line() +
    ggtitle('Age vs Probability of death')
```



Q1(a): Use `death_prob` to determine the death probability of 50 year old female.

```
death_prob %>% filter(age == 50 & sex =='Female')
```

```
##   age    sex     prob
## 1  50 Female 0.003193
```

Q1(b): What is the Expected Value(E) for a 50 year old female

```r
p <- 0.003193
loss_payout*p + profit_premium*(1-p)
```

```
## [1] 667.3781
```

Q1(c): Calculate the Standard Error (SE) for the same.

```r
sqrt(p*(1-p))*abs(loss_payout - profit_premium)
```

```
## [1] 8527.332
```

Q1(d): (E) for all 1000 policies for 50 year old females

```r
En <- n*(loss_payout*p + profit_premium*(1-p))
En
```

```
## [1] 667378.1
```

Q1(e): (SE) for all 1000 policies for 50 year old females.

```r
SEn <- sqrt(n*p*(1-p))*abs(loss_payout - profit_premium)
SEn
```

```
## [1] 269657.9
```

Q1(f): Using Central Limit Theorem to calculate the probability of losing money on this set of 1000 policies.

```r
pnorm(0, En, SEn)
```

```
## [1] 0.006663556
```

Section 2: Lets perform similar tasks for 50 year old Male.

Q2(a): Probability of death for a 50 year old Male

```r
prob <- death_prob %>% filter(age == 50 & sex == 'Male')
prob
```

```
##   age  sex     prob
## 1  50 Male 0.005013
```

Q2(b): Suppose from 1000 policies the expected profit is to be $700000. What should be the premium?

```
# En = n * (loss*p + premium*(1-p))
# premium = (En - n*loss*p)/(n*(1-p))
En <- 700000
p <- prob$prob
premium <- (700000 - n*loss_payout*p)/(n*(1 - p))
premium
```

```
## [1] 1459.265
```

Q2(c): calculate the SE for n policies with previous premium amount.

```
SEn <- sqrt(n*p*(1-p))*abs(loss_payout-premium)
SEn
```

```
## [1] 338262.1
```

Q2(d): Using Central Limit Theorem what is the probability of losing money on 1000 policies.

```
pnorm(0, En, SEn)
```

```
## [1] 0.01925424
```

Section 3: An event changes the death_prob, what will be change in profit for the insurance company.

```
new_p <- 0.015
loss_payout <- -150000
premium <- 1150
n <- 1000
```

Q3(a): What is the new expected value?

```
# E = n*(loss*new_p + premium*(1-new_p))
En <- n*(loss_payout*new_p + premium*(1-new_p))
En
```

```
## [1] -1117250
```

Q3(b): What is the new expected value?

```
# SE = sqrt(n*new_p*(1-new_p))*abs(loss-premium)
SEn <- sqrt(n*new_p*(1-new_p))*abs(loss_payout-premium)
SEn
```

```
## [1] 580994.3
```

Q3(c): What is the probability of company losing money.

```
pnorm(0, En, SEn)
```

```
## [1] 0.9727597
```

Q3(d): What is the probability of losing more than 1 million $?

```
pnorm(-1000000, En, SEn)
```

```
## [1] 0.5799671
```

Q3(e): New death probability are as defined below, what is the lowest death probability for which chance of losing money exceed 90%?

```
p <- seq(.01, .03, .001)
En <- n*(p*loss_payout + premium*(1-p))
SEn <- sqrt(n*p*(1-p))*abs(loss_payout - premium)
prob_losing_money <- pnorm(0, En, SEn)
df <- data.frame(prob = p, losing_money = prob_losing_money)
df %>% filter(losing_money>0.9)
```

```
##     prob losing_money
## 1  0.013    0.9338629
## 2  0.014    0.9573137
## 3  0.015    0.9727597
## 4  0.016    0.9827809
## 5  0.017    0.9892027
## 6  0.018    0.9932761
## 7  0.019    0.9958377
## 8  0.020    0.9974369
## 9  0.021    0.9984289
## 10 0.022    0.9990410
## 11 0.023    0.9994167
## 12 0.024    0.9996465
## 13 0.025    0.9997864
## 14 0.026    0.9998713
```

```
## 15 0.027     0.9999226
## 16 0.028     0.9999536
## 17 0.029     0.9999723
## 18 0.030     0.9999834
```

Q3(f): For the new death probabilities what is the least death probability for which the chance to lose over 1million$ exceeds 90%?

```
p <- seq(.01, .03, .0025)
En <- n*(p*loss_payout + premium*(1-p))
SEn <- sqrt(n*p*(1-p))*abs(loss_payout - premium)
prob_losing_money <- pnorm(-1000000, En, SEn)
df <- data.frame(prob = p, losing_money = prob_losing_money)
df %>% filter(losing_money>0.9)
```

```
##     prob losing_money
## 1 0.0200    0.9039858
## 2 0.0225    0.9611879
## 3 0.0250    0.9854673
## 4 0.0275    0.9948727
## 5 0.0300    0.9982746
```

Q4(a): Define a sampling model for n = 1000 loans, p_loss = 0.015, loss = -150000, $premium = 1150$. What is the reported profit in millions?

Set the seed to 25 using following command: set.seed(25, sample.kind = "Rounding").

```
set.seed(25, sample.kind = "Rounding")
```

```
n <- 1000
p_loss <- 0.015
loss <- 150000
premium <- 1150
report <- sample(c(-1, 1), n, replace = TRUE,
                          prob = c(p_loss, 1-p_loss))
profit <- 1150*sum(report == 1)/10^6
loss <- 150000*sum(report == -1)/10^6
profit - loss
```

```
## [1] -1.41955
```

Q4(b): set the seed to 27, and run the simulation with 10000 replicates. Finally, find out what is the probability of losing 1 million $ or more?

```
set.seed(27, sample.kind = 'Rounding')
```

```
## Warning in set.seed(27, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```r
B <- 10000
profits <- replicate(B, {
    report <- sample(c(-1, 1), n, replace = TRUE,
                     prob = c(p_loss, 1 - p_loss))
    profit <- 1150*sum(report == 1)/10^6
    loss <- 150000*sum(report == -1)/10^6
    total <- profit - loss
})
mean(profits <= -1)
```

```
## [1] 0.5388
```

Q5(a): For a death prob of 0.015, the probability of losing money should be less than 5%. Calculate the premium required.

Pr(S < 0) = 0.05

Pr((S - En)/SEn < -En/SEn) = 0.05

Pr(Z < - En/SEn) = 0.05

therefore, -En/SEn = qnorm(0.05)

```r
# -[(loss*p + premium*(1-p))*n/(sqrt(n*p*(1-p))*abs(loss-premium)) = qnorm(0.05)
loss_payout <- -150000
p <- 0.015
n <- 1000
z <- qnorm(0.05)

# -(-2250000 + 985*premium)/(3.843*premium + 576570) = -1.6448
# 2250000 - 985*premium = -6.320966*premium - 948342
# premium = (2250000 + 948342)/(985 - 6.320966)
premium <- (2250000 + 948342)/(985 - 6.30966)
premium
```

```
## [1] 3267.982
```

Q5(b): Expected profit per policy at above premium rate?

```r
En <- loss_payout*p + premium*(1-p)
En
```

```
## [1] 968.9619
```

Q5(c): What is the expected profit for over 1000 policies?

```r
En*1000
```

```
## [1] 968961.9
```

Q5(d): Using Monte Carlo Simulation determine the probability of losing money for 10000 trials.

```
set.seed(28, sample.kind = "Rounding")
```

```
## Warning in set.seed(28, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
B <- 10000
n <- 1000
loss_payout <- -150000
premium <- 3267.982
p <- 0.015

profits <- replicate(B, {
    draws <- sample(c(loss_payout, premium), n, replace = TRUE,
                    prob = c(p, 1-p))
    sum(draws)
})
mean(profits < 0)
```

```
## [1] 0.0554
```

Section 6: The pandemic changes the probability of payout by a value between -0.01 to 0.01, using this changes calculate the profit returns.

Set seed to 29.

```
set.seed(29, sample.kind = "Rounding")
```

```
## Warning in set.seed(29, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

Q6(a): What is the expected value for 1000 policies?

```
B <- 10000
n <- 1000
loss_payout <- -150000
premium <- 3267.982

profits <- replicate(B, {
    p <- 0.015 + sample(seq(-0.01, 0.01, length = 100), 1)
    draws <- sample(c(loss_payout, premium), n, replace = TRUE,
                    prob = c(p, 1-p))
    sum(draws)
})
mean(profits)
```

```
## [1] 968226.6
```

Q6(b): What is the probability of losing money?

```r
mean(profits < 0)
```

```
## [1] 0.1908
```

Q6(c): What is the probability of losing more than 1 million $?

```r
mean(profits < -1000000)
```

```
## [1] 0.0424
```