STAT2201, Semester 1, 2017

Solution for Assignment 2

Questions marked for grade (each 20%): Q2, Q3, Q5, Q8

Question 1 - A Discrete Distribution - PMF

Errata Notes:

- In item (d), 21 can also be 2.1. The solution below is for 21, but solutions that treat it as 2.1 are also valid.

Consider the function $p_{XY}(\cdot,\cdot)$:

x	y	$p_{XY}(x,y)$
1.0	1.0	1/4
1.5	2.0	1/8
1.5	3.0	1/4
2.5	4.0	1/4
3.0	5.0	1/8

Determine the following:

(a) Show that $p_{X,Y}$ is a valid probability mass function.

(b) $P(X < 2.5,\ Y < 3)$.

(c) $P(X < 2.5)$.

(d) $P(Y < 3)$.

(e) $P(X > 1.8,\ Y > 4.7)$.

(f) $E(X),\ E(Y),\ V(X),\ V(Y)$.

(g) Are X and Y independent random variables?

(h) $P(X + Y \le 4)$.

Solution:

(a) $p(1,1)+p(1.5,2)+p(1.5,3)+p(2.5,4)+p(3,5)=1$and $p(x,y)\geq 0$ for all x,y.

(b) $P(X<2.5,Y<3)=p(1,1)+p(1.5,2)=3/8$

(c) $P(X<2.5)=p(1,1)+p(1.5,2)+p(1.5,3)=5/8$

(d) $P(Y<3)=p(1,1)+p(1.5,2)=3/8$

(e) $P(X>1.8,Y>4.7)=p(3,5)=1/8$

(f)
$E(X)=1*P(X=1)+1.5*P(X=1.5)+2.5*P(X=2.5)+3*P(X=3)=1*p(1,1)+$
$*\,p(2.5,4)+3*p(3,5)=1*p(1,1)+1.5*(p(1.5,2)+p(1.5,3))+2.5*p(2.5,4)+3*p(3,5)$
$+\,3*(1/8)=1.8125$

$E(Y)=1*P(Y=1)+2*P(X=2)+3*P(X=3)+4*P(X=4)+5*P(X=5)=1*p$
$*\,p(2.5,4)+5*p(3,5)=1*(1/4)+2*(1/8)+3*(1/4)+4*(1/4)+5*(1/8)=2.875$

$E(X^2)=1^2*(1/4)+1.5^2*(3/8)+2.5^2*(1/4)+3^2*(1/8)=3.78125$

$E(Y^2)=1^2*(1/4)+2^2*(1/8)+3^2*(1/4)+4^2*(1/4)+5^2*(1/8)=10.125$

$V(X)=E(X^2)-(E(X))^2=3.78125-1.8125^2=0.496$

$V(Y)=E(Y^2)-(E(Y))^2=10.125-2.875^2=1.859$

(g) No, the random varianbles are not independent. Knowing the value of X gives information about the value of Y and vice-versa.

(h) $P(X+Y\leq 4)=p(1,1)+p(1.5,2)=3/8$

Question 2 - A Discrete Distribution - CDF

Given the cdf, $F(x)$ below for the random variable, X, calculate the following:

$$F(x) = \begin{cases} 0 & x < -10 \\ 0.25 & -10 \leqslant x < 30 \\ 0.75 & 30 \leqslant x < 50 \\ 1 & 50 \leqslant x \end{cases}$$

(a) $P(X \leqslant 50)$

(b) $P(X \leqslant 40)$

(c) $P(40 \leqslant X \leqslant 60)$

(d) $P(X < 0)$

(e) $P(0 \leqslant X < 10)$

(f) $P(-10 < X < 10)$

(g) The mean.

(h) The standard deviation.

(i) Sketch the probability mass function (PMF).

Solution:

(a) $P(X \leq 50) = F(50) = 1$

(b) $P(X \leq 40) = F(40) = 0.75$

(c) $P(40 \leq X \leq 60) = F(60) - F(40) = 1 - 0.75 = 0.25$

(d) $P(X < 0) = F(0^-) = 0$

Note that x^- implies the x minus an minute quantity. This is because "<" is used instead of "$\leq$"

(e) $P(0 \leq X < 10) = F(10^-) - F(0) = 0 - 0 = 0$

(f) $P(-10 < X < 10) = F(10^-) - F(-10^-) = 0 - 0 = 0$

(g) It is best To first answer (i) and see the PMF.

$\mu = -10p(-10) + 30p(30) + 50p(50) = -10 * 0.25 + 30 * 0.5 + 50 * 0.25 = 25$

(h)Compute the second moment: \$\mu_2 = (-10)^2 *0.25* + *30*^2 0.5 + 50^2 0.25 = 1100

Now variance = \$\mu_2 - mu^2 = 1100 - 25^2 = 475

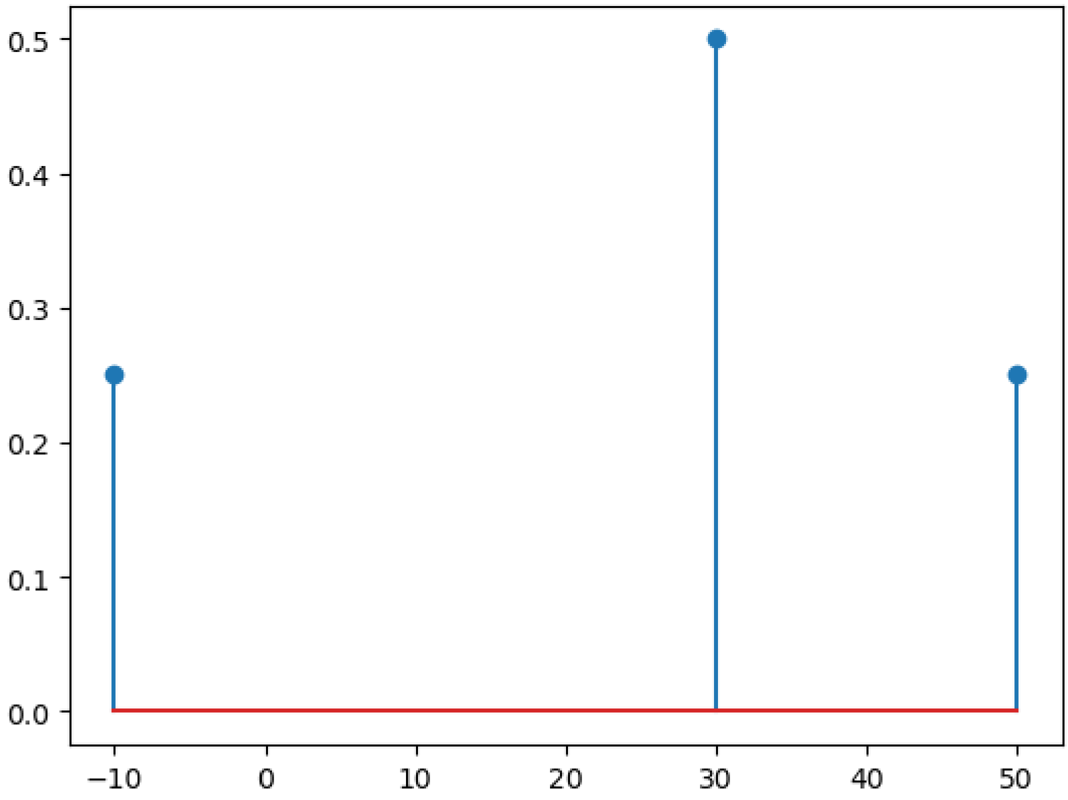
Now standard deviation is the square root of the variance: $\sqrt{475} = 21.79$

(i) The support (range) of the distribution is $\{-10, 30, 50\}$. The pmf is:

$p(-10) = 0.25$, $p(30) = 0.5$, $p(50) = 0.25$

In [6]:

```
using PyPlot
x = [-10, 30, 50]
px = [0.25, 0.5, 0.25]
PyPlot.stem(x,px);
```



Question 3 - Guessing on Multiple Choice Exams

A multiple-choice test contains 25 questions, each with four answers. Assume that a student just guesses on each question.

(a) What is the probability that the student answers more than 20 questions correctly?

(b) What is the probability that the student answers fewer than 5 questions correctly?

The following code generates the vector pmfValues from a Binomial distribution with parameters $n = 10$ and $p = 0.4$. It then sums up the vector, illustrating that the sum of all of the probabilities is 1.

```
using Distributions
bDist = Binomial(10,0.4)
pmfValues = [pdf(bDist, x) for x in 0:20]
sum(pmfValues[1:11])
```

(c) Modify the code above, to validate your answers in (a) and (b).

Solution:

The number of correct questions is $X \sim Bin(25, 0.25)$

(a) $P(X > 20) = \sum_{i=21}^{25} \binom{25}{i} \, 0.25^i \, 0.75^{25-i} \approx 0$

(b) $P(X < 5) = \sum_{i=0}^{4} \binom{25}{i} \, 0.25^i \, 0.75^{25-i} = 0.2138$

(c) The code below outputs a tuple with (answer to a, answer to b)

In [9]:

```
using Distributions
bDist = Binomial(25 ,0.25)
pmfValues = [pdf(bDist , x) for x in 0:25]
sum(pmfValues[22:26]),sum(pmfValues[1:5])
```

Out[9]:

```
(1.2434600904498447e-8,0.21374092343881484)
```

Note that the support of the Binmoial distribution is 0:25 (26 values in total), but arrays in Julia re indexed from 1 onwards (in this case 1...26). Hence when accessing the pmfValues in the array a shift of 1 takes place.

Question 4 - Stuck in Traffic

A particularly long traffic light on your morning commute is green 20% of the time that you approach it. Assume that each morning represents an independent trial.

(a) Over 5 mornings, what is the probability that the light is green on exactly one day?

(b) Over 20 mornings, what is the probability that the light is green on exactly four days?

(c) Over 20 mornings, what is the probability that the light is green on more than four days?

(d) What is the mean number of days with green light during a month of 30 days?

Optional: You may verify your analytic answers using Julia, in a similar manner to Question 3c.

In [9]:

```
30*0.2
```

Out[9]:

```
6.0
```

Solution:

If the number of mornings is n, the the number of greens is a Binomial random variable, $N \sim Bin(n, 0.2)$.

(a) Here $n = 5$: $P(N = 1) = \binom{5}{1} 0.2^1 0.8^4 = 5 * 0.2 * 0.8^4 = 0.4096$

(b) Here $n = 20$: $P(N = 4) = \binom{20}{4} 0.2^4 0.8^1 6 = \frac{20*19*18*17}{4*3*2} * 0.2^4 * 0.8^{16} = 0.218$

(c) Again $n = 20$:
$P(N > 4) = 1 - P(N \leq 4) = 1 - P(N = 4) - P(N = 3) - P(N = 2) - P(N = 1) - P(N = ($
$- \binom{20}{2} 0.2^2 0.8^{18} - \binom{20}{1} 0.20.8^{19} - 0.8^{20} = 0.3703$

(d) Here $n = 30$: $E(N) = 30 * 0.2 = 6$

In [13]:

```
using Distributions
pdf(Binomial(5,0.2),1), pdf(Binomial(20,0.2),4), 1-cdf(Binomial(20,0.2),4)
```

Out[13]:

```
(0.4096,0.21819940194610052,0.37035173609733096)
```

Question 5 - Aerospace Inspections

The thickness of a flange on an aircraft component is Uniformly distributed between 0.95 and 1.05 millimetres. Determine the following:

(a) Cumulative distribution function of flange thickness.

(b) Proportion of flanges that exceeds 1.02 millimetres.

(c) Thickness exceeds 90% of the flanges.

(d) Mean and variance of flange thickness.

(e) Assume now that you are sampling 10 independent flanges. What is the variance of the number of flanges with a thickness less than 0.97 millimetres?

Solution:

(a) In a uniform distribution, $f(x) = K$ for $x \in [0.05, 1.05]$ with $f(x) = 0$ otherwise. Now since $1 = \int_{x=0.95}^{1.05} K\,dx$, it holds that $K = 10$.

Now the CDF, $F(x) = 0$ for $x < 0.95$ and $F(x) = 1$ for $x > 1.05$. Then inbetween it holds that,

$F(x) = \int_{u=0.95}^{x} 10\,du = 0.1\,u\,\big|_{u=0.95}^{u=x} = 10(x - 0.95)$.

(b) $P(X > 1.02) = 1 - P(X \leq 1.02) = 1 - F(1.02) = 1 - 10(1.02 - 0.95) = 0.3$.

(c) This is the 0.9 quantile. Solve: $0.9 = F(x)$ or

$0.9 = 10(x - 0.95)$.

The solution is $x = 1.04$.

(d) You can compute the mean directly by $\int x\, f(x)dx$ (and similarly for the variance). But here is an alternative method:

Set $U \sim \text{uniform}(0, 1)$. Then $X = 0.1U + 0.95$.

For U we know that $E[U] = 1/2$ and $Var(U) = 1/12$.

Now $E[X] = 0.1E[U] + 0.95 = 1.0$.

Further, $Var(X) = 0.1^2 Var(U) = 0.01/12 = .0008333$.

(e) Let N be the number of flanges with a thinkness in the desired level. We have,

$p = F(0.97) = 0.2$ as the chance of being in the desired level.

And $N \sim \text{Bin}(10, p)$.

So $Var(N) = p(1-p) * N = 0.2 * 0.8 * 10 = 1.6$.

Question 6 - Mobile Phone Semiconductors

The line width for semiconductor manufacturing is assumed to be Normally distributed with a mean of 0.5 micrometers and a standard deviation of 0.05 micrometers.

(a) What is the probability that a line width is greater than 0.62 micrometer?

(b) What is the probability that a line width is between 0.47 and 0.63 micrometer?

(c) The line width of 90% of samples is below what value?

Solution:

$X \sim N(0.5, 0.05^2)$

(a) $P(X > 0.62) = P(\frac{X-0.5}{0.05} > \frac{0.62-0.5}{0.05}) = P(Z > 2.4) = 1 - \Phi(2.4) = 0.0082$

(b)
$P(0.47 < X < 0.63) = P(-0.6 < Z < 2.6) = \Phi(2.6) - \Phi(-0.6) = 0.99534 - 0.274253 = 0.72$

(c) $x_{0.9}$: $0.9 = P(X < x_{0.9}) = P(\frac{X-0.5}{0.05} < \frac{x_{0.9}-0.5}{0.05}) = P(Z < \frac{x_{0.9}-0.5}{0.05}) = \Phi(\frac{x_{0.9}-0.5}{0.05})$,

Hence, $\Phi^{-1}(0.9) = \frac{x_{0.9}-0.5}{0.05}$ or $0.05 * \Phi^{-1}(0.9) + 0.5 = x_{0.9}$

$\Phi^{-1}(0.9) = 1.282$

Hence, $x_{0.9} = 0.5641$. Here it is with Julia:

In [25]:

```
quantile(Normal(0.5,0.05),0.9)
```

Out[25]:

```
0.56407757827723
```

In [26]:

```
quantile(Normal(),0.9)
```

Out[26]:

```
1.2815515655446006
```

Question 7 - The Prototype Shoe

The weight of a sophisticated running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.5 ounce.

(a) What is the probability that a shoe weighs more than 13 ounces?

(b) What must the standard deviation of weights be in order for the company to state that 99% of its shoes weighs less than 13 ounces?

(c) If the standard deviation remains at 0.5 ounce, what must the mean weight be for the company to state that 99% of its shoes weighs less than 13 ounces?

Solution:

(a) Now more quickly than previous exercise:
$P(Z > (13 - 12)/0.5) = P(Z > 2) = 1 - \Phi(2) = 0.02275$

(b) $X \sim N(12, \sigma^2$. Want $0.99 = P(X < 13)$

or $0.99 = P(Z < (13 - 12)/\sigma) = \Phi((13 - 12)/\sigma)$

or $\Phi^{-1}(0.99) = 2.326 = (13 - 12)/\sigma$

Hence $\sigma = 1/2.326 = 0.43$

(c) in same manner:

$2.326 = (13 - \mu)/0.5$ Hence $\mu = 13 - 2.326 * 0.5 = 11.837$

In [30]:

```
#Check (b)
cdf(Normal(12,0.43),13)
```

Out[30]:

```
0.9899795534792678
```

In [33]:

```
#Check (c)
cdf(Normal(11.837,0.5),13)
```

Out[33]:

```
0.9899907246591324
```

Question 8 - Time Until (Blue Screen of Death) BSoD

Suppose that the time to failure (in hours) of hard drives in a personal computer can be modelled by an exponential distribution with $\lambda = 0.0003$.

(a) What proportion of the hard drives will last at least 10,000 hours?

(b) What proportion of the hard drives will last at most 7,000 hours?

(c) What is the variance of the time until failure for a hard drive?

(d) Use Monte Carlo simulation to predict the following: Assume a computer now has two independent hard-drives and the failure of the computer occurs once both hard-drives have died. What is the mean life of the computer?

Solution:

(a) For $X \sim \exp(\lambda)$, $P(X > x) = 1 - F(x) = e^{-\lambda x}$. Hence,

$e^{-0.0003*10^4} = e^{-3} = 0.04979$

(b) Now asking $P(X < 7000) = 1 - e^{-0.0003*7000} = 0.8775$

(c) A variance of $exp(\lambda)$ random variables is $1/\lambda^2$ so, $Var(X) \approx 1.11E7$

(d) Here the lifetime of a computer is the maximum of two independent exponential random variables. Our simulation doesn't use the distributions package but rather uses the fact that $-\frac{1}{\lambda}Log(U)$ is distributied like $exp(\lambda)$ when U is uniform(0,1).

In [23]:

```
meanLifeEstimate =  mean([max((-1/0.0003)*log(rand()),(-1/0.0003)*log(rand())) for _ in
 1:10^6])
```

Out[23]:

```
5006.363358082411
```