Named Continuous Distribution Application Solutions

Lt Col Ken Horton   Professor Bradley Warner

14 September, 2020

Exercises

For problems 1-3 below, ***1)*** define a random variable that will help you answer the question, ***2)*** state the distribution and parameters of that random variable; ***3)*** determine the expected value and variance of that random variable, and ***4)*** use that random variable to answer the question.

1. On a given Saturday, suppose vehicles arrive at the USAFA North Gate according to a Poisson process at a rate of 40 arrivals per hour.

    a) Find the probability no vehicles arrive in 10 minutes.

X: number of vehicles that arrive in 10 minutes

$X \sim \textsf{Pois}(\lambda = 40/6 = 6.67)$ and $\mathrm{E}(X) = \mathrm{Var}(X) = 6.67$.

$\mathrm{P}(\text{no arrivals in 10 minutes}) = \mathrm{P}(X = 0) = \frac{6.67^0 e^{-6.67}}{0!} = e^{-6.67}$

```
exp(-40/6)
```

```
## [1] 0.001272634
```

```
##or
dpois(0,40/6)
```

```
## [1] 0.001272634
```

or, using the exponential distribution:

Y: time in minutes until the next arrival

$Y \sim \textsf{Expon}(\lambda = 40/60 = 0.667)$ and $\mathrm{E}(Y) = 1.5$ and $\mathrm{Var}(Y) = 2.25$.

$$\mathrm{P}(\text{at least 10 minutes until the next arrival}) = \mathrm{P}(Y \geq 10) = \int_{10}^{\infty} \frac{2}{3} e^{-\frac{2}{3}y} \, \mathrm{d}y$$

```
1-pexp(10,2/3)
```

```
## [1] 0.001272634
```

or using simulation:

```r
set.seed(616)
mean(rpois(100000,40/6) == 0)
```

```
## [1] 0.00126
```

```r
mean(rexp(100000,2/3) >=10)
```

```
## [1] 0.00127
```

b) Find the probability that at least 5 minutes will pass before the next arrival.

Y: same as in part a

$$\text{P(at least 5 minutes until next arrival)} = \text{P}(Y \geq 5) = \int_5^\infty \frac{2}{3} e^{-\frac{2}{3}y}\,\text{d}y$$

```r
1-pexp(5,2/3)
```

```
## [1] 0.03567399
```

c) Find the probability that the next vehicle will arrive between 2 and 10 minutes from now.

Same Y as defined above.

```r
pexp(10,2/3)-pexp(2,2/3)
```

```
## [1] 0.2623245
```

d) Find the probability that at least 7 minutes will pass before the next arrival, given that 2 minute have already passed. Compare this answer to part (b). This is an example of the memoryless property of the exponential distribution.

$$\text{P}(Y \geq 7 | Y \geq 2) = \frac{\text{P}(Y \geq 7, Y \geq 2)}{\text{P}(Y \geq 2)} = \frac{\text{P}(Y \geq 7)}{\text{P}(Y \geq 2)}$$

```r
(1-pexp(7,2/3))/(1-pexp(2,2/3))
```

```
## [1] 0.03567399
```

This is the same answer and a result of the memoryless property.

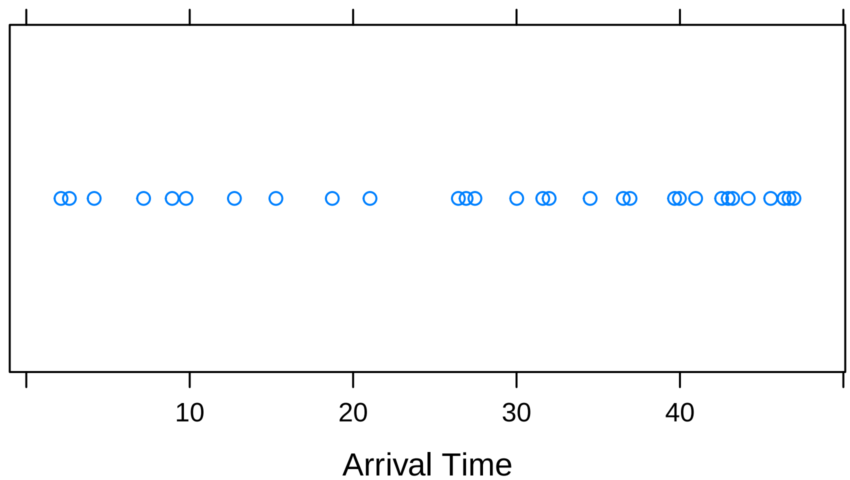
e) Fill in the blank. There is a probability of 90% that the next vehicle will arrive within ___ minutes. This value is known as the 90% percentile of the random variable.

```r
qexp(0.9,2/3)
```

```
## [1] 3.453878
```

f) Use the function `stripplot()` to visualize the arrival of 30 vehicles using a random sample from the appropriate exponential distribution.

```r
set.seed(202)
stripplot(cumsum(rexp(30,2/3)),xlab="Arrival Time")
```



2. Suppose time until computer errors on the F-35 follows a Gamma distribution with mean 20 hours and variance 10.

a) Find the probability that 20 hours pass without a computer error.

X: time in hours until next computer error.

$X \sim \mathsf{Gamma}(\alpha = 40, \lambda = 2)$

We need to find α and λ from the given moments.

$\mathrm{E}(X) = 20 = \frac{\alpha}{\lambda}$

$\mathrm{Var}(X) = 10 = \frac{\alpha}{\lambda^2}$

Notice that $\frac{\mathrm{E}(X)}{\mathrm{Var}(X)} = \lambda = \frac{20}{10} = 2$ and then using $\mathrm{E}(X) = 20 = \frac{\alpha}{\lambda}$ we get $\alpha = 40$.

$\mathrm{P}(X \geq 20)$:

```
1-pgamma(20,shape=40,rate=2)
```

```
## [1] 0.4789711
```

b) Find the probability that 45 hours pass without a computer error, given that 25 hours have already passed. Does the memoryless property apply to the Gamma distribution?

$$P(X \geq 45 | X \geq 25) = \frac{P(X \geq 45, X \geq 25)}{P(X \geq 25)} = \frac{P(X \geq 45)}{P(X \geq 25)}$$

```
(1-pgamma(45,40,2))/(1-pgamma(25,40,2))
```

```
## [1] 1.77803e-08
```

No, the memoryless property does not apply to the Gamma distribution.

c) Find a and b where there is a 95% probability that the time until next computer error will be between a and b. (Note: technically, there are many answers to this question, but find a and b such that each tail has equal probability.)

```
qgamma(c(0.025,0.975),40,2)
```

```
## [1] 14.28829 26.65714
```

So in the time interval $[14, 29, 26.66]$.

```
qgamma(.95,40,2)
```

```
## [1] 25.46987
```

Another answer is between $[0, 25, 47]$.

3. Suppose PFT scores in the cadet wing follow a normal distribution with mean 330 and standard deviation 50.

a) Find the probability a randomly selected cadet has a PFT score higher than 450.

X: PFT score of a randomly selected cadet

$X \sim \textsf{Norm}(\mu = 330, \sigma = 50)$

$\text{E}(X) = 330$ and $\text{Var}(X) = 50^2 = 2500$.

```
1-pnorm(450,330,50)
```

```
## [1] 0.008197536
```

b) Find the probability a randomly selected cadet has a PFT score within 2 standard deviations of the mean.

Need $\text{P}(230 \leq X \leq 430)$.

```
pnorm(430,330,50)-pnorm(230,330,50)
```

```
## [1] 0.9544997
```

c) Find a and b such that 90% of PFT scores will be between a and b.

Need a such that $\text{P}(X \leq a) = 0.05$ and b such that $\text{P}(X \geq b) = 0.05$:

```
qnorm(0.05,330,50)
```

```
## [1] 247.7573
```

```
qnorm(0.95,330,50)
```

```
## [1] 412.2427
```

d) Find the probability a randomly selected cadet has a PFT score higher than 450 given he/she is among the top 10% of cadets.

Need $\text{P}(X > 450 | X > x_{0.9})$ where $x_{0.9}$ is the 90th percentile of X.

The 90th percentile is:

```
qnorm(0.9,330,50)
```

```
## [1] 394.0776
```

$$\text{P}(X > 450 | X > x_{0.9}) = \frac{\text{P}(X > 450, X > x_{0.9})}{\text{P}(X > x_{0.9})} = \frac{\text{P}(X > 450, X > 394.08)}{\text{P}(X > x_{0.9})} = \frac{\text{P}(X > 450)}{0.1}$$

This is assuming that $x_{0.9} < 450$. Otherwise the problem is trivial and the probability is 1.

```r
(1-pnorm(450,330,50))/0.1
```

```
## [1] 0.08197536
```

4. Let $X \sim \mathsf{Beta}(\alpha = 1, \beta = 1)$. Show that $X \sim \mathsf{Unif}(0, 1)$. Hint: write out the beta distribution pdf where $\alpha = 1$ and $\beta = 1$.

The beta pdf is:

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

When $X \sim \mathsf{Beta}(\alpha = 1, \beta = 1)$, this becomes:

$$f_X(x) = \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^{1-1}(1-x)^{1-1} = 1$$

5. When using `R` to calculate probabilities related to the gamma distribution, we often use `pgamma`. Recall that `pgamma` is equivalent to the cdf of the gamma distribution. If $X \sim \mathsf{Gamma}(\alpha, \lambda)$, then

$$\mathrm{P}(X \leq x) = \mathsf{pgamma(x,alpha,lambda)}$$

The `dgamma` function exists in `R` too. In plain language, explain what `dgamma` returns. I'm not looking for the definition found in `R` documentation. I'm looking for a simple description of what that function returns. Is the output of `dgamma` useful? If so, how?

The `dgamma` function returns the value of probability density function. While this is not a probability, it is still a useful quantity. It can be said that larger densities ($f(x)$) imply that values near x are more likely to occur than values associated with smaller densities. It is also useful when computing conditional probability distributions.

6. Advanced. You may have heard of the 68-95-99.7 rule. This is a helpful rule of thumb that says if a population has a normal distribution, then 68% of the data will be within one standard deviation of the mean, 95% of the data will be within two standard deviations and 99.7% of the data will be within three standard deviations. Create a function in `R` that has two inputs (a mean and a standard deviation). It should return a vector with three elements: the probability that a randomly selected observation from the normal distribution with the inputted mean and standard deviation lies within one, two and three standard deviations. Test this function with several values of mu and sd. You should get the same answer each time.

```
rulethumb<-function(mu,sd){
  pnorm(mu+c(1,2,3)*sd,mu,sd)-pnorm(mu-c(1,2,3)*sd,mu,sd)
}

rulethumb(15,12)
```

```
## [1] 0.6826895 0.9544997 0.9973002
```

7. Derive the mean of a general uniform distribution, $U(a,b)$.

From the definition

$$E(X)=\int_a^b xf(x)dx=$$

$$=\int_a^b \frac{x}{b-a}dx=$$

$$=\frac{1}{b-a}\int_a^b xdx=\frac{1}{b-a}\cdot\frac{x^2}{2}\bigg|_a^b=$$

$$=\frac{1}{b-a}\cdot\frac{b^2-a^2}{2}=\frac{1}{b-a}\cdot\frac{(b-a)(b+a)}{2}=\frac{(b+a)}{2}$$

File creation information

- File creation date: 2020-09-14
- Windows version: Windows 10 x64 (build 18362)
- R version 3.6.3 (2020-02-29)
- `mosaic` package version: 1.7.0
- `tidyverse` package version: ‘r