STAT 400 Homework 06

Spring 2018 | Dalpiaz | UIUC

Due: *Friday, March 9, 2:00 PM*

Exercise 1

Consider a random variable X with the moment generating function

$$M_X(t) = e^{5t+8t^2} = \exp(5t + 8t^2)$$

(a) Calculate $P(4 < X < 16)$.

Solution:

Since the moment generating function for a normal random variable X, with mean μ and variance σ^2 is given by

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

we know that the moment generating function given is that of a normal random variable with

$$\mu = 5$$

and

$$\frac{\sigma^2}{2} = 8$$

Thus we have

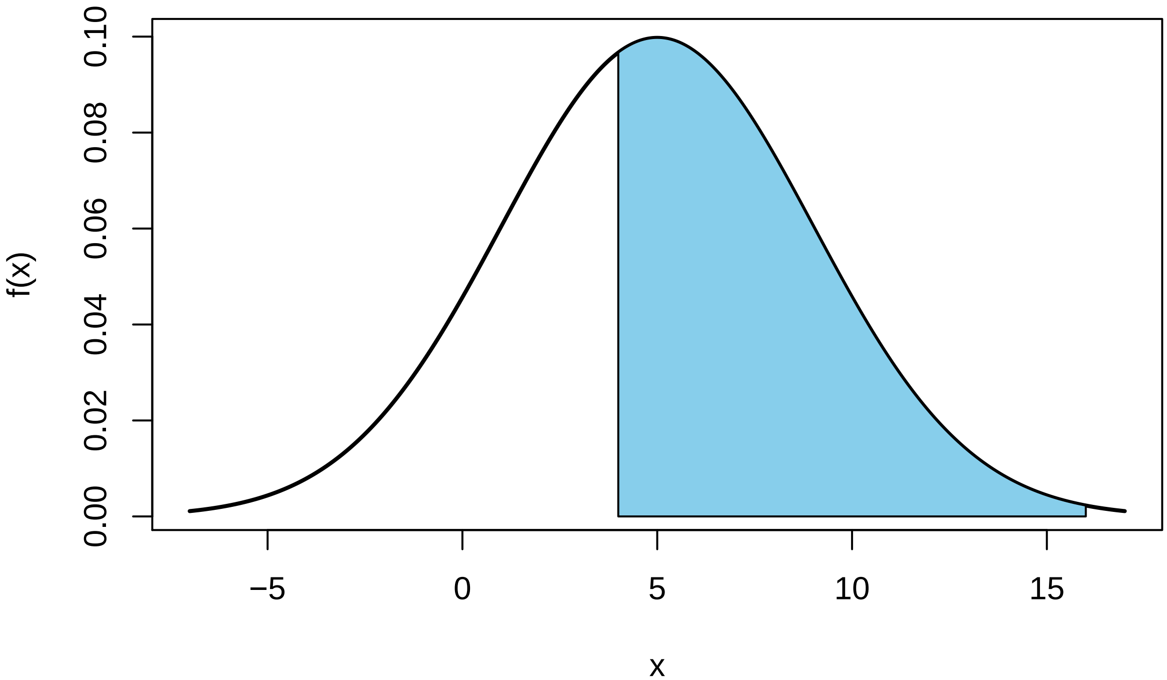
$$\text{E}[X] = 5$$

and

$$\text{Var}[X] = 16.$$

Then

$$\begin{aligned} P(4 < X < 16) &= P\left(\frac{4-5}{4} < Z < \frac{16-5}{4}\right) = P(-0.25 < Z < 2.75) \\ &= P(Z < 2.75) - P(Z < -0.25) = 0.9970 - 0.4013 = \boxed{0.5957} \end{aligned}$$




```r
c(pnorm(-0.25), pnorm(2.75))
```

```
## [1] 0.4012937 0.9970202
```

```r
(pnorm(2.75) - pnorm(-0.25))
```

```
## [1] 0.5957266
```

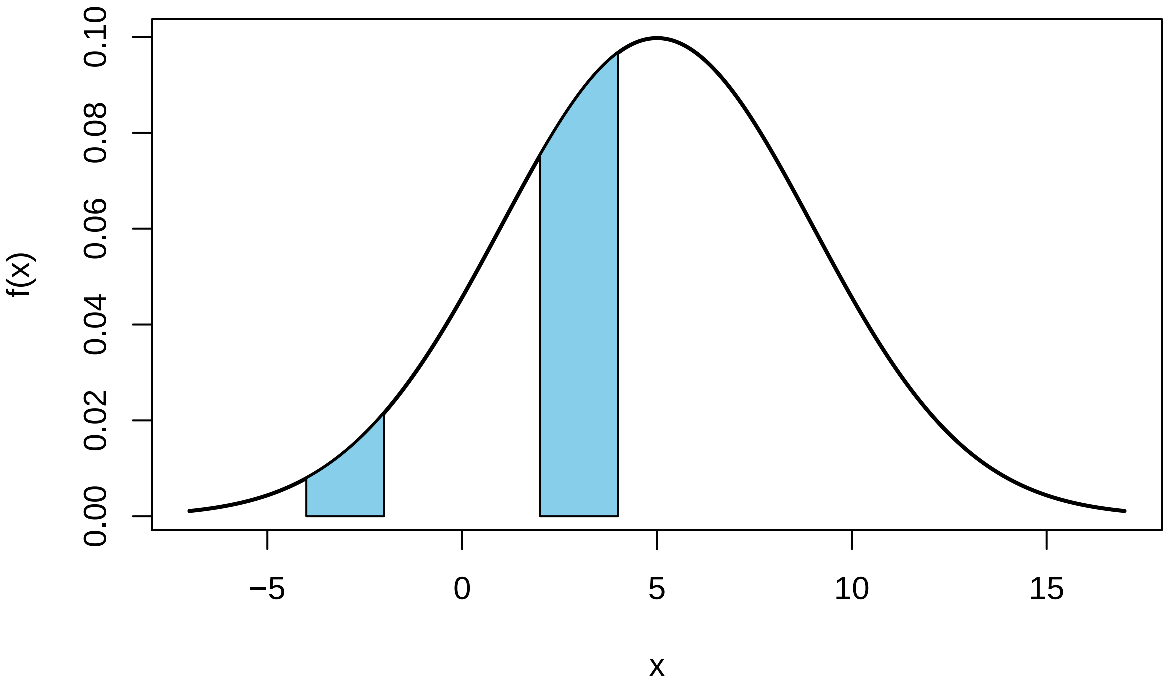
```r
diff(pnorm(c(4, 16), mean = 5, sd = 4))
```

```
## [1] 0.5957266
```

(b) Calculate $P(4 < X^2 < 16)$.

Solution:

$$
\begin{aligned}
P(4 < X^2 < 16) &= P(-4 < X < -2) + P(2 < X < 4) \\
&= P(-2.25 < Z < -1.75) + P(-0.75 < Z < -0.25) \\
&= (0.0401 - 0.0122) + (0.4013 - 0.2266) = \boxed{0.2026}
\end{aligned}
$$




```r
c(pnorm(-2.25), pnorm(-1.75), pnorm(-0.75), pnorm(-0.25))
```

```
## [1] 0.01222447 0.04005916 0.22662735 0.40129367
```

```r
(pnorm(-1.75) - pnorm(-2.25)) + (pnorm(-0.25) - pnorm(-0.75))
```

```
## [1] 0.202501
```

```r
diff(pnorm(c(-4, -2), mean = 5, sd = 4)) +
  diff(pnorm(c(2, 4), mean = 5, sd = 4))
```

```
## [1] 0.202501
```

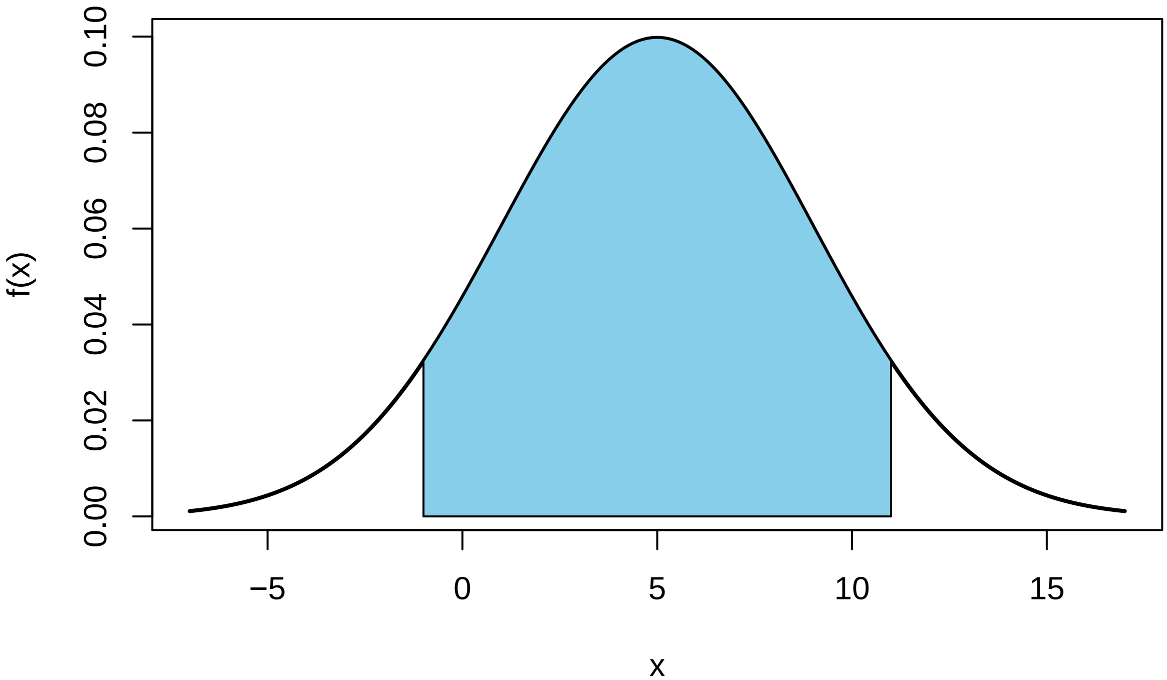
Exercise 2

Consider a random variable X with $\text{E}[X] = 5$ and $\text{Var}[X] = 16$.

(a) Calculate $P(|x - 5| < 6)$ if X follows a normal distribution.

Solution:

$$\begin{aligned} P(|X-5| \le 6) &= P[|X-5| \le (1.5) \cdot (4)] \\ &= P(-1.5 < Z < 1.5) \\ &= P(Z < -1.5) - P(Z < 1.5) \\ &= 0.9332 - 0.0668 \\ &= \boxed{\mathbf{0.8664}} \end{aligned}$$




```r
diff(pnorm(c(-1, 11), mean = 5, sd = 4))
```

```
## [1] 0.8663856
```

(b) Use Chebyshev's inequality to provide a lower bound for $P(|x-5|<6)$. (No longer assume X is normal.)

Solution:

$$
\begin{aligned}
P(|X-5| \leq 6) &= P(|X-5| \leq (1.5)\cdot(4)) \\
&\geq 1 - \frac{1}{(1.5)^2} \\
&= \boxed{\mathbf{0.5556}}
\end{aligned}
$$

Exercise 3

In the original Pokémon Red and Blue, there were 151 Pokémon, but only 150 of these Pokémon could actually be caught or obtained in the games. If you wanted to "catch 'em all", the 151st Pokémon, Mew, could only be obtained through special giveaway events at local video game retailers. (For example FuncoLand, which was later purchased by GameStop.)

Suppose that these giveaway events at your local FuncoLand occur according to a Poisson process with an average of one event per two months.

For this exercise, assume that all 12 months of the year have the same number of days. Additionally, assume that the four seasons all last exactly three months.

- Winter: December, January, February
- Spring: March, April, May
- Summer: June, July, August
- Fall: September, October, November

Also, suppose that it is the beginning of a new year, that is January 1, 1999.

(a) What is the probability that the first event occurs before Spring?

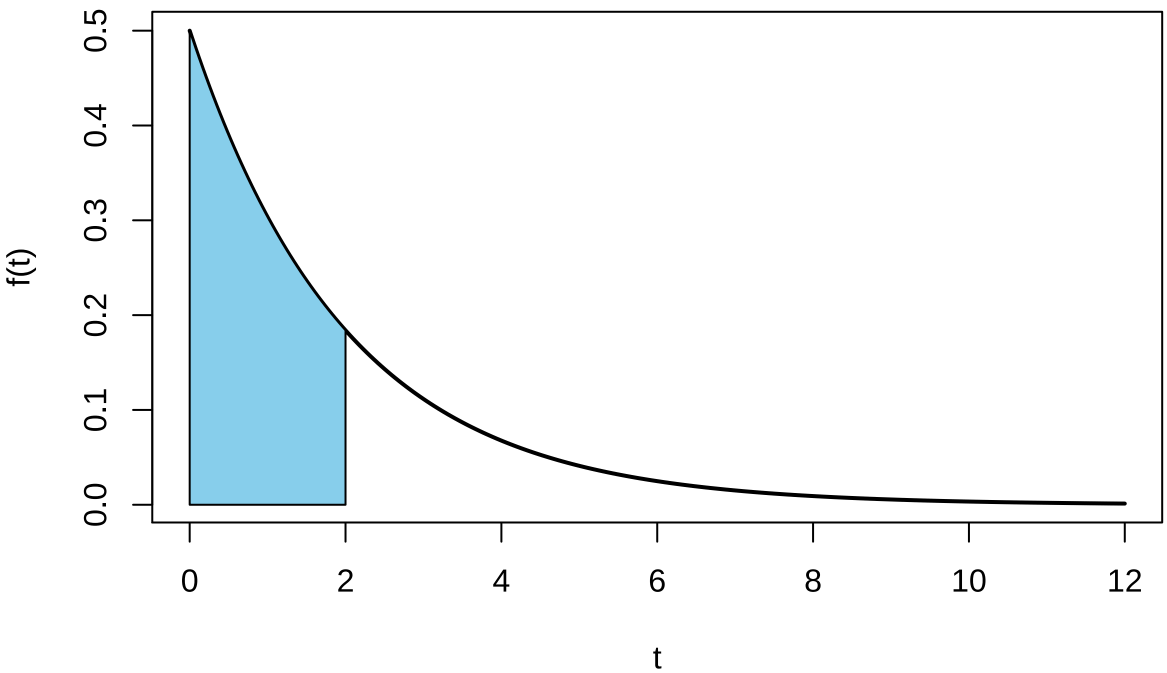
Solution:

Define:

- T_k as the waiting time (in months) until the kth event
- X_t as the number of events in time (months) t

Here, T_1 follows an exponential distribution with $\lambda = 0.5$.

$$P(T_1 < 2) = 1 - e^{-(0.5)(2)} = \boxed{0.6321}$$




```r
pexp(2, rate = 0.5)
```

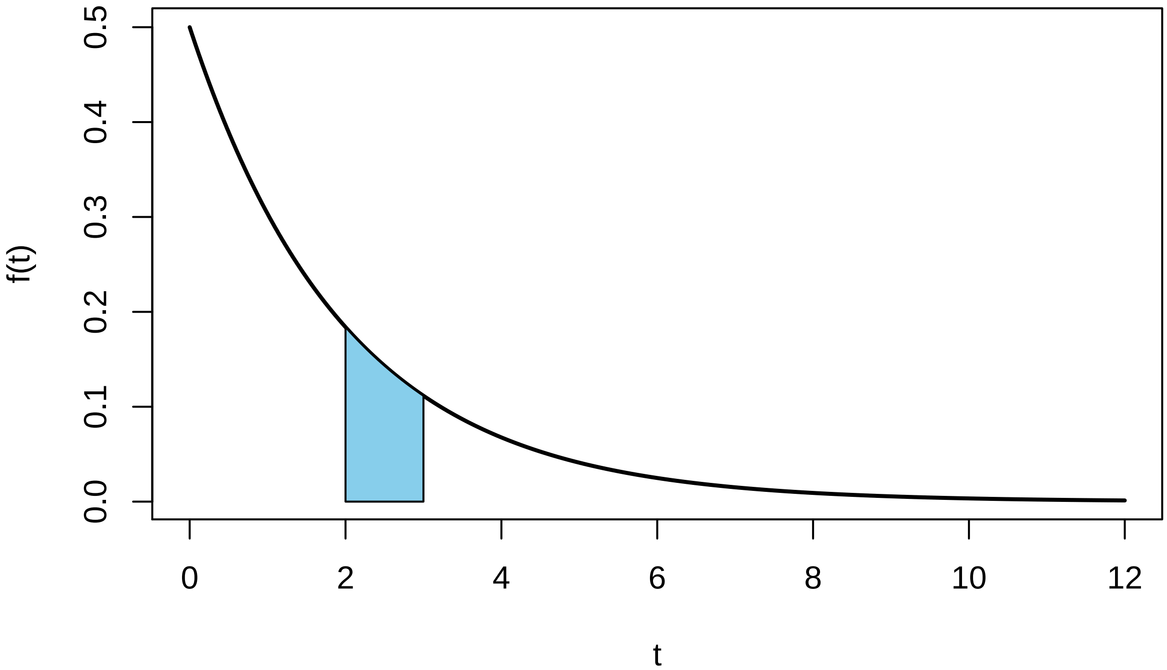
```
## [1] 0.6321206
```

(b) What is the probability that the first event occurs during the month of March?

Solution:

Again, T_1 follows an exponential distribution with $\lambda = 0.5$.

$$P(2 < T_1 < 3) = P(T_1 > 2) - P(T_1 > 3) = e^{-(0.5)(2)} - e^{-(0.5)(3)} = \boxed{0.1447}$$




```r
pexp(2, rate = 0.5, lower.tail = FALSE) -
  pexp(3, rate = 0.5, lower.tail = FALSE)
```

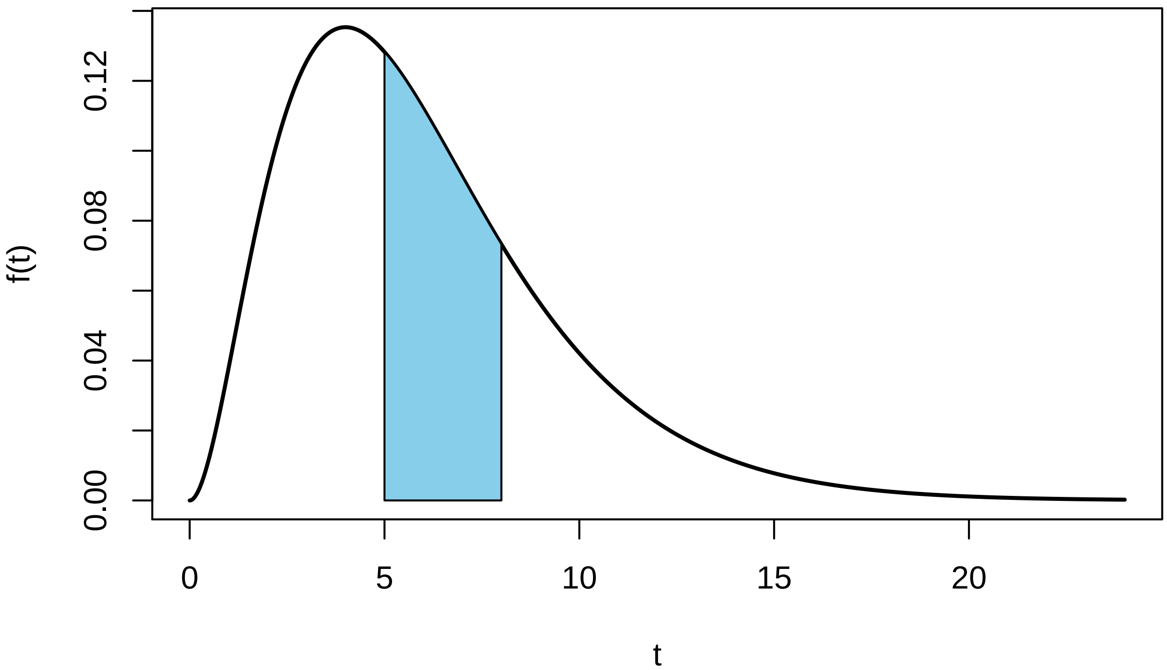
```
## [1] 0.1447493
```

(c) What is the probability that the third event occurs during Summer?

Solution:

- T_3 is the waiting time (in months) until the 3rd event
    - $T_3 \sim \text{Gamma}(\alpha = 3, \lambda = 0.5)$
- X_5 is the number of events in 5 months
    - $X_5 \sim \text{Pois}(\lambda_5 = 2.5)$
- X_8 is the number of events in 8 months
    - $X_8 \sim \text{Pois}(\lambda_8 = 4)$

$$
\begin{aligned}
P(5 < T_3 < 8) &= P(T_3 > 5) - P(T_3 > 8) \\
&= P(X_5 \leq 2) - P(X_8 \leq 2) \\
&= 0.5438 - 0.2381 \\
&= \boxed{0.3057}
\end{aligned}
$$




```r
ppois(2, lambda = 2.5) - ppois(2, lambda = 4)
```

```
## [1] 0.3057098
```

```r
diff(pgamma(c(5, 8), shape = 3, rate = 0.5))
```

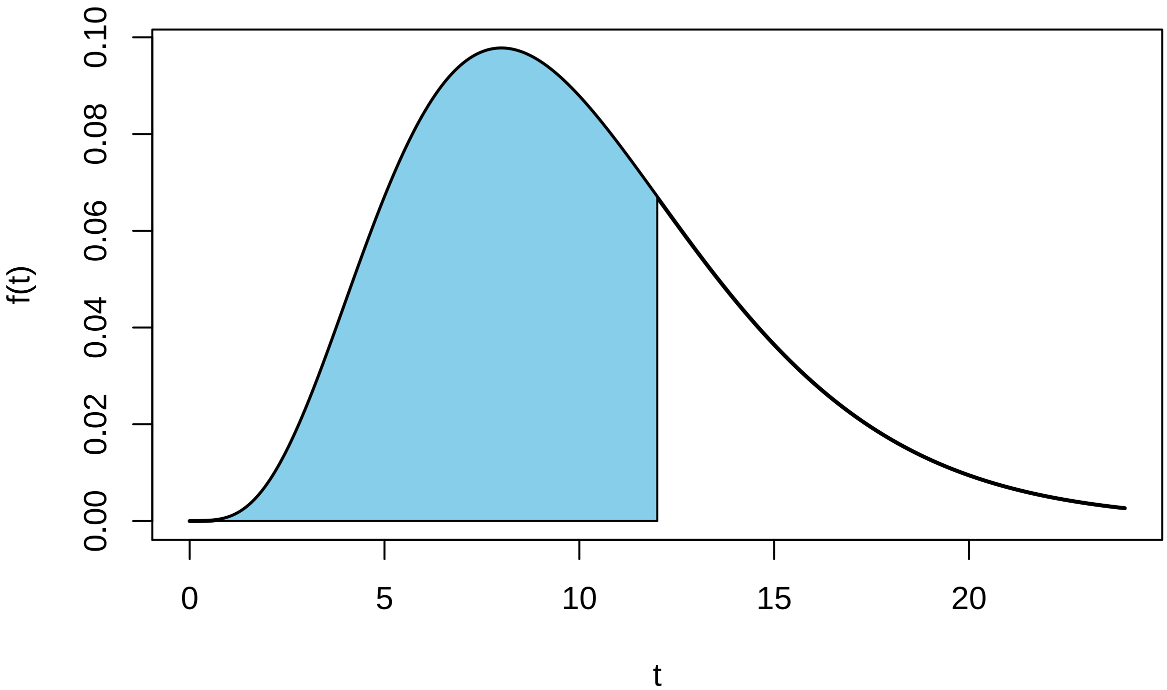
```
## [1] 0.3057098
```

(d) What is the probability that the fifth event occurs before the end of the year?

Solution:

- T_5 is the waiting time (in months) until the 5th event
  - $T_5 \sim \text{Gamma}(\alpha = 5, \lambda = 0.5)$
- X_{12} is the number of events in 12 months
  - $X_{12} \sim \text{Pois}(\lambda_{12} = 6)$

$$P[T_5 < 12] = 1 - P[T_5 > 12] = 1 - P[X_{12} \leq 4] = \boxed{0.7149}$$




```
1 - ppois(4, lambda = 6)
```

```
## [1] 0.7149435
```

```
pgamma(12, shape = 5, rate = 0.5)
```

```
## [1] 0.7149435
```

Exercise 4

In Neverland, annual income (in \$), X, is distributed according to a Gamma distribution with $\alpha = 5$ and $\theta = 10,000$. Every year, the IRS audits 1% of the individuals with an income below \$50,000, 3% of individuals with incomes between \$50,000 and \$95,000, and 6% of individuals with an income above \$95,000. Suppose that the individuals to be audited are selected at random.

(a) What is the distribution of the income groups? That is, what proportion of Neverland's population falls into each of the three income groups? Even more specifically, find $P(X < \$50,000)$, $P(\$50,000 < X < \$95,000)$, and $P(X > \$95,000)$.

- Hint: These probabilities should add to 1.

Solution:

Recall that if $T \sim \text{Gamma}(\alpha = 5, \theta = \frac{1}{\lambda})$ and α is an integer, then

$$P(T > t) = P(Y \leq \alpha - 1)$$

where $Y_{\lambda t} \sim \text{Pois}(\lambda t = \frac{t}{\theta})$

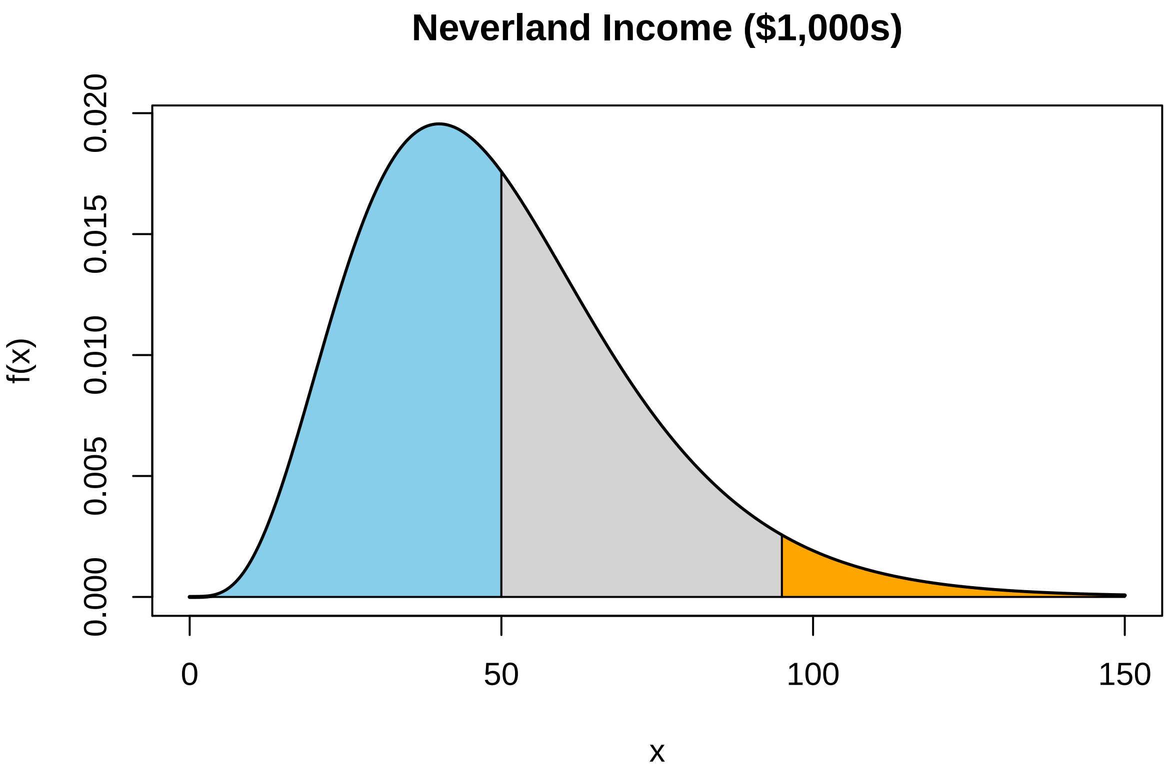
Here $X \sim \text{Gamma}(\alpha = 5, \theta = 10,000)$. Then we also have $\lambda = \frac{1}{10,000}$.

We first obtain some necessary probabilities.

- $P(X > 50,000) = P(Y_5 \leq 4) = 0.440$
- $P(X > 95,000) = P(Y_{9.5} \leq 4) = 0.040$

We then use these to obtain the desired distribution.

- $P(X < 50,000) = 1 - 0.440 = \boxed{0.56}$
- $P(50,000 < X < 95,000) = 0.440 - 0.040 = \boxed{0.40}$
- $P(X > 95,000) = \boxed{0.04}$




```r
pgamma(50000, shape = 5, scale = 10000)
```

```
## [1] 0.5595067
```

```r
diff(pgamma(c(50000, 95000), shape = 5, scale = 10000))
```

```
## [1] 0.4002306
```

```r
pgamma(95000, shape = 5, scale = 10000, lower.tail = FALSE)
```

```
## [1] 0.04026268
```

(b) You overhear Mr. Statman complain about being audited. What is the probability that Mr. Statman's income is below $\$50,000$? Between $\$50,000$ and $\$95,000$? Above $\$95,000$?

- Hint: Again, these probabilities should add to 1. Essentially, we're finding the (posterior) distribution of Mr. Statman's possible income (group) given that he's being audited.

Solution:

We are given the following information about audits.

- $P(A \mid X < 50,000) = 0.01$
- $P(A \mid 50,000 < X < 95,000) = 0.03$
- $P(A \mid X > 95,000) = 0.06$

We use this information, as well as the previous calculated distribution to obtain some intermediate probabilities.

- $P(A \cap X < 50,000) = (0.01) \cdot (0.56) = 0.0056$
- $P(A \cap 50,000 < X < 95,000) = (0.03) \cdot (0.40) = 0.03$
- $P(A \cap X > 95,000) = (0.06) \cdot (0.04) = 0.06$

We use these probabilities to obtain the overall probability of an audit, ignoring income.

$$P(A) = 0.0056 + 0.03 + 0.06 = 0.02$$

We then repeatedly apply conditional probability (Bayes) to obtain the desired distribution.

$$P(X < 50,000 \mid A) = \frac{0.0056}{0.02} = \boxed{0.28}$$
$$P(50,000 < X < 95,000 \mid A) = \frac{0.03}{0.02} = \boxed{0.60}$$
$$P(X > 95,000 \mid A) = \frac{0.06}{0.02} = \boxed{0.12}$$

Notice that these probabilities do add to one, as this is a conditional **distribution**.