

Problem Set 1

1. Let X, X_1, \dots, X_n be independent, exponentially distributed random variables with parameter λ .

(a) Find the median of X , i.e. a number m such that $\mathbb{P}(X < m) = \frac{1}{2}$, in terms of λ .

To find the median m of $X \sim \exp(\lambda)$, compute:

$$\begin{aligned}\frac{1}{2} &= \mathbb{P}(X \leq m) \\ &= \int_0^m f_X(x) dx \\ &= \int_0^m \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_{x=0}^{x=m} \\ &= 1 - e^{-\lambda m}.\end{aligned}$$

Solving $\frac{1}{2} = 1 - e^{-\lambda m}$, we have that the median of $X \sim \exp(\lambda)$ is $m = \frac{\ln(2)}{\lambda}$.

- (b) Let the random variable $M := \text{median}(X_1, \dots, X_n)$. Usually computation of medians must take into account the possibility of tied values; e.g. see [this site](#). But not here. Why not?

By nature of continuous random variables, the possibility that for any two $j, k \in \{1, \dots, n\}$ be such that $X_j = X_k = c$, for any $c \in [0, \infty)$ is zero. This is because:

$$\mathbb{P}(X_j = X_k) = \int_0^\infty \int_c^c \lambda_j e^{-\lambda x_j} \lambda_k e^{-\lambda x_k} dx_j dx_k = \int_0^\infty 0 dx_k = 0.$$

Having more than two ties has the same zero probability, because that probability is naturally supposed to be no greater than the previously calculated possibility.

Now if we expand the chances of any number of tied values, this becomes a (finite) combinatorial problem, which culminates to:

$$\begin{aligned}\mathbb{P}(\text{Any number of tied values in } X_1, \dots, X_n) &= \sum_{i=2}^n \binom{n}{i} \mathbb{P}(\text{Any } i \text{ ties in } X_1, \dots, X_n) \\ &= \sum_{i=2}^n 0 \\ &= 0.\end{aligned}$$

(c) Find $\mathbb{P}\left(X_1 < \frac{6}{5}X_2\right)$.

The region for $(x_1, x_2) \in (\mathbb{R}_{\geq 0})^2$ defined $x_1 < \frac{6}{5}x_2$ is the region

$$D := \left\{ (x_1, x_2) \in (\mathbb{R}_{\geq 0})^2 : x_1 < \frac{6}{5}x_2 \text{ for all } x_2 \in \mathbb{R} \right\}.$$

Rewrite $\mathbb{P}\left(X_1 < \frac{6}{5}X_2\right)$ as $\mathbb{P}\left(\frac{6}{5}X_2 - X_1 > 0\right)$. Then for $X_1, X_2 \sim \exp(\lambda)$:

$$\begin{aligned} \mathbb{P}\left(\frac{6}{5}X_2 - X_1 > 0\right) &= \int_D f_{X_1, X_2}(x_1, x_2) d(x_1, x_2) \\ &= \int_D f_{X_1}(x_1) f_{X_2}(x_2) d(x_1, x_2), && \text{as } X_1 \text{ and } X_2 \text{ are indep.;} \\ &= \int_0^\infty \int_0^{\frac{6}{5}x_2} (\lambda e^{-\lambda x_1}) (\lambda e^{-\lambda x_2}) dx_1 dx_2 \\ &= \int_0^\infty \left[\int_0^{\frac{6}{5}x_2} \lambda e^{-\lambda x_1} dx_1 \right] \lambda e^{-\lambda x_2} dx_2 \\ &= \int_0^\infty \left[1 - e^{-\frac{6}{5}\lambda x_2} \right] \lambda e^{-\lambda x_2} dx_2 \\ &= \int_0^\infty f_{X_2}(x_2) dx_2 - \int_0^\infty f_{X_2}\left(\frac{11x_2}{5}\right) dx_2 \\ &= \int_0^\infty f_{X_2}(x_2) dx_2 - \frac{5}{11} \int_0^\infty f_{X_2}(u) du, && \text{where } u := \frac{11}{5}x_2, dx = \frac{5}{11}du; \\ &= 1 - \frac{5}{11} \cdot 1 \\ &= \frac{6}{11}. \end{aligned}$$

- (d) Say $n = 2$, and that X_j is exponentially distributed with parameter λ_j , $j = 1, 2$. We choose one of the X_j at random, with probability p_j . Let T be the result. Find $\mathbb{E}[T]$ and $\text{Var}(T)$ in terms of the λ_j and the p_j .

Since $n = 2$, we have that T is a Bernoulli probability on $X_j \sim \exp(\lambda_j)$ with a probability of p_j , $j = 1, 2$. To keep in format with this problem's preferred answer format, we will not call $p_2 = 1 - p_1$, and simply refer to each of them as p_j .

By the **law of total expectation**,

$$\begin{aligned}\mathbb{E}[T] &= \sum_{j=1}^2 \mathbb{P}(T = X_j) \mathbb{E}[T \mid T = X_j] \\ &= \sum_{j=1}^2 p_j \mathbb{E}[X_j] \\ &= \sum_{j=1}^2 p_j \frac{1}{\lambda_j},\end{aligned}\quad \text{since } X_j \sim \exp(\lambda_j) \text{ implies } \mathbb{E}[X_j] = \frac{1}{\lambda_j};$$

and,

$$\begin{aligned}\mathbb{E}[T^2] &= \sum_{j=1}^2 \mathbb{P}(T = X_j) \mathbb{E}[T^2 \mid T = X_j] \\ &= \sum_{j=1}^2 p_j \mathbb{E}[X_j^2] \\ &= \sum_{j=1}^2 p_j \int_0^\infty x_j f_{X_j}(x_j) dx_j \\ &= \sum_{j=1}^2 p_j \int_0^\infty \lambda_j x_j^2 e^{-\lambda_j x_j} dx_j, \\ &= \sum_{j=1}^2 p_j \lim_{t_j \rightarrow \infty} \left[- \left(x_j^2 + \frac{2}{\lambda_j} x_j + \frac{2}{\lambda_j^2} \right) e^{-\lambda_j x_j} \right]_{x_j=0}^{x_j=t_j} \\ &= \sum_{j=1}^2 p_j \lim_{t_j \rightarrow \infty} \left[\frac{2}{\lambda_j^2} - \left(t_j^2 + \frac{2}{\lambda_j} t_j + \frac{2}{\lambda_j^2} \right) e^{-\lambda_j t_j} \right] \\ &= \sum_{j=1}^2 \frac{2p_j}{\lambda_j^2},\end{aligned}\quad \text{using L'Hôpital's rule.}$$

Then $\mathbb{E}[T] = \frac{p_1}{\lambda_1} + \frac{p_2}{\lambda_2}$, and by the **law of total variance**,

$$\text{Var}(T) = \mathbb{E}[T^2] - \mathbb{E}[T]^2 = \sum_{j=1}^2 \frac{2p_j}{\lambda_j^2} - \left(\sum_{j=1}^2 \frac{p_j}{\lambda_j} \right)^2 = \frac{2p_1 - p_1^2}{\lambda_1^2} - 2 \frac{p_1 p_2}{\lambda_1 \lambda_2} + \frac{2p_2 - p_2^2}{\lambda_2^2}.$$

2. Say the random variable X has an exponential distribution with parameter λ . However, our observation of it is truncated at c , so that our observed value $T := \min(X, c)$. Derive a formula for $\text{Var}(T)$ as a function of λ and c , using the “**Pythagorean Theorem for variance**” (**Law of Total Variance**, SRC p.54). Write simulation code to verify your formula.

The **law of total variance** tells us that $\text{Var}(T) = \mathbb{E}[T^2] - \mathbb{E}[T]^2$. We calculate each of these values, using the **law of total expectation** on T^k for $k = 1, 2$:

$$\mathbb{E}[T^k] := \mathbb{E}[T^k \mid X \leq c] \mathbb{P}(X \leq c) + \mathbb{E}[T^k \mid X > c] \mathbb{P}(X > c).$$

However, notice that $\mathbb{E}[T^k \mid X > c] = c^k$, since the expectation is conceptually the most likely value of T^k , which if $X > c$ automatically makes $T^k \equiv c^k$, exactly. (One may also make a uniform distribution argument to show that $\mathbb{E}[T^k \mid X > c] = c^k$.)

Also, $X \sim \exp(\lambda)$ implies $\mathbb{P}(X > c) = \int_c^\infty f_X(x) dx = e^{-\lambda c}$, and $\mathbb{P}(X \leq c) = 1 - \mathbb{P}(X > c) = 1 - e^{-\lambda c}$.

This leaves:

$$\begin{aligned} \mathbb{E}[T \mid X \leq c] &:= \int_0^c \lambda x e^{-\lambda x} dx & \mathbb{E}[T^2 \mid X \leq c] &:= \int_0^c \lambda x^2 e^{-\lambda x} dx \\ &= -x e^{-\lambda x} \Big|_{x=0}^{x=c} + \int_0^c e^{-\lambda x} dx & &= -x^2 e^{-\lambda x} \Big|_{x=0}^{x=c} + \frac{2}{\lambda} \mathbb{E}[T \mid X \leq c] \\ &= -x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \Big|_{x=0}^{x=c} & &= -c^2 e^{-\lambda c} + \frac{2}{\lambda} \left(-c - \frac{1}{\lambda} \right) e^{-\lambda c} \\ &= \frac{1}{\lambda} - \left(c + \frac{1}{\lambda} \right) e^{-\lambda c} & &= \frac{2}{\lambda^2} - \left(c^2 + \frac{2c}{\lambda} + \frac{2}{\lambda^2} \right) e^{-\lambda c}. \end{aligned}$$

Thus,

$$\begin{aligned} \text{Var}(T)(c, \lambda) &= \mathbb{E}[T^2] - \mathbb{E}[T]^2 \\ &= \mathbb{P}(X \leq c) \mathbb{E}[T^2 \mid X \leq c] + \mathbb{P}(X > c) \mathbb{E}[T^2 \mid X > c] \\ &\quad - (\mathbb{P}(X \leq c) \mathbb{E}[T \mid X \leq c] + \mathbb{P}(X > c) \mathbb{E}[T \mid X > c])^2 \\ &= (1 - e^{-\lambda c}) \int_0^c \lambda x^2 e^{-\lambda x} dx + c^2 e^{-\lambda c} - \left((1 - e^{-\lambda c}) \int_0^c \lambda x e^{-\lambda x} dx + c e^{-\lambda c} \right)^2 \end{aligned}$$

Conceptually speaking, the fact that T is constant whenever $X \geq c$ implies that $\text{Var}(T) \leq \text{Var}(X)$. As an cursory check, we see that, for example $\text{Var}(T)(1, 3) = 0.07950529 \leq 0.111\dots = 1/3^2 = \text{Var}(X)$.

We checked that our random variable is correctly written, by checking R’s internal variance calculation, and saw a very strong resemblance for large sample sizes.

```
>print(diff_n)
[1] 0.0246611975 0.0088572821 0.0049489390 0.0009356232
```

It is not generally consistent throughout every simulation, but the general trend is that the difference of the variance of 10^n sample using R’s calculations, versus our mathematical computation for $\text{Var}(T) \approx 0.07950529$ which represents the variance of N samples as $N \rightarrow \infty$ gets smaller, for $n = 1, 2, 3, 4$.

So we decided do the average of one-hundred trial differences between each 10^n data sample experiments, and $\text{Var}(T) \approx 0.07950529$, for $n = 1, \dots, 5$.

```
> print(avg_diff_n)
[1] 0.031991824 0.008758512 0.003163073 0.002013893 0.001828636
```

This clearly exhibits a trend wherein higher sample sizes correlates to lower difference between variance values of the data, versus the mathematical calculation of $\text{Var}(T)$.

Therefore the $\text{Var}(T)(c, \lambda)$, where $c = 1$ and $\lambda = 3$ is ≈ 0.07950529 over N sample sizes as $N \rightarrow \infty$, as desired.

See `Problem2.R` for the program script.

3. The **Tower Property** says,

$$\mathbb{E} [\mathbb{E}[Y \mid U, V] \mid U] = \mathbb{E}[Y \mid U]$$

Write simulation code that demonstrates this for a context in which $\mathbb{E}[Y \mid U, V] = U + V$, $\mathbb{E}[Y \mid U] = U + \mathbb{E}[V \mid U]$.

See `Problem3.R` for the program script.

Following the hint in blog, we use the following Bernoulli probabilities: **A** is defined as 0 or 1; **B** is defined as 0 or 10 and **C** is defined as 0 or 100; each with probability of **X=0** of 3/4, 1/3, 1/2, and **W** is -1 or 1 with probabilities 1/2.

Then we defined `U <- A + C`, `V <- B + C`, and `Y <- U + V + W`.

Then we ran the simulations to experimentally verify that $\mathbb{E}[Y \mid U, V] = U + V$, $\mathbb{E}[Y \mid U] = U + \mathbb{E}[V \mid U]$.

Running the simulation on our `Problem3.R` file, using 10000 samples for a single experiment.

Running more samples will give us a smaller discrepancy between the left-hand sides and right-hand sides of each equation.