STAT 515

Homework #5 WITH SOLUTIONS

This homework must be submitted electronically to ANGEL. I strongly encourage the use of LaTeX.

1. Let X_1 and X_2 be independent exponential random variables with rates λ_1 and λ_2, respectively. Let

$$X_{(1)} = \min\{X_1, X_2\} \quad \text{and} \quad X_{(2)} = \max\{X_1, X_2\}.$$

We have shown in class that $X_{(1)}$ is exponential with rate $\lambda_1 + \lambda_2$.

(a) Find $EX_{(2)}$. (**Hint:** What is $E[X_{(1)} + X_{(2)}]$?)

Solution: Since $X_{(1)} + X_{(2)} = X_1 + X_2$, we conclude that

$$EX_{(2)} = E(X_1 + X_2) - E(X_{(1)}) = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}.$$

(b) Find a probability density function for $X_{(2)}$ and use it to calculate $\text{Var } X_{(2)}$.

Solution: Start with the cdf for $X_{(2)}$:

$$F_{X_{(2)}}(x) = P(\max\{X_1, X_2\} \le x) = P(X_1 \le x)P(X_2 \le x) = (1 - e^{-\lambda_1 x})(1 - e^{-\lambda_2 x})$$

Now differentiate to get a density function:

$$f_{X_{(2)}}(x) = \frac{d}{dx}(1 - e^{-\lambda_1 x} - e^{-\lambda_2 x} + e^{-(\lambda_1+\lambda_2)x}) = \lambda_1 e^{-\lambda_1 x} + \lambda_2 e^{-\lambda_2 x} - (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)x}.$$

Interestingly, this is just the sum of two exponential densities minus a third! We can use this fact to find $E(X_{(2)}^2)$ quickly because we know that for an exponential random variable Y with rate μ, $E(Y^2) = \text{Var}\,(Y) + [E(Y)]^2 = 2/\mu^2$:

$$E(X_{(2)}^2) = \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} - \frac{2}{(\lambda_1 + \lambda_2)^2}.$$

We conclude that

$$\text{Var } X_{(2)} = E(X_{(2)}^2) - [EX_{(2)}]^2 = \frac{2}{\lambda_1^2} + \frac{2}{\lambda_2^2} - \frac{2}{(\lambda_1 + \lambda_2)^2} - \left[\frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}\right]^2.$$

After simplification, we get

$$\text{Var } X_{(2)} = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} - \frac{3}{(\lambda_1 + \lambda_2)^2}.$$

(c) Find $\text{Cov}\,(X_{(1)}, X_{(2)})$. (**Hint:** What is $\text{Var}\,[X_{(1)} + X_{(2)}]$?)

Solution: We know that $2\,\text{Cov}\,(X_{(1)}, X_{(2)}) = \text{Var}\,[X_{(1)} + X_{(2)}] - \text{Var } X_{(1)} - \text{Var } X_{(2)}$. Furthermore, $\text{Var}\,[X_{(1)} + X_{(2)}]$ is simply $\text{Var } X_1 + \text{Var } X_2$ since X_1 and X_2 are independent and they have the same sum as $X_{(1)}$ and $X_{(2)}$. Thus,

$$\text{Cov}\,(X_{(1)}, X_{(2)}) = \frac{1}{2}\left(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} - \frac{1}{(\lambda_1 + \lambda_2)^2} - \text{Var } X_{(2)}\right) = \frac{1}{(\lambda_1 + \lambda_2)^2}.$$

Actually, it is just as easy to find this covariance directly, since $E(X_{(1)}X_{(2)}) = E(X_1 X_2) = E(X_1)E(X_2)$:

$$\text{Cov}\,(X_{(1)}, X_{(2)}) = E(X_{(1)}X_{(2)}) - E(X_{(1)})E(X_{(2)}) = \frac{1}{\lambda_1 \lambda_2} - \frac{1}{\lambda_1 + \lambda_2}\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}\right) = \frac{1}{(\lambda_1 + \lambda_2)^2}.$$

This also gives us a way to check that the answer for (b) is correct (!), since we could use this result as an alternative method of finding $\text{Var } X_{(2)}$.

2. Theorem 5.2 in Section 5.3.5 states that in a Poisson process $N(t)$ with rate λ, given that $N(t) = n$, the n arrival times $S_1, \ldots, S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $(0, t)$, i.e.,

$$P(S_1 = t_1, \ldots, S_n = t_n \mid N(t) = n) = \frac{n!}{t^n} I(0 < t_1 < \cdots < t_n).$$

(a) Clearly describe the general algorithm this suggests for simulating a Poisson process on an interval $[0, t]$. (**Hint**: you will simulate the process in two stages.)

Solution: First, generate $N \sim \text{Poisson}(\lambda t)$, then generate N i.i.d. Uniform$(0, t)$ variables, sort them, and take $S_1, \ldots, S_N$ to be the sorted values.

(b) Consider a *homogeneous* Poisson process with $\lambda = 10$. Using the algorithm from part (a), simulate 10,000 realizations of the above Poisson process on the interval $[0, 5]$.

Solution: The trick here is storing a different number of arrival times for each realization. One way to do it using an R object called a list:

```
> X <- list()  # Use double-brackets to refer to list items, e.g., X[[1]]
> for (i in 1:10000) {
+   X[[i]] <- sort(runif(rpois(1, 50), min=0, max=5))
+ }
```

(c) Report the sample mean for the number of events in the interval (0,1) and the number of events in the interval (4,5). How do these means compare with the corresponding theoretical expectations?

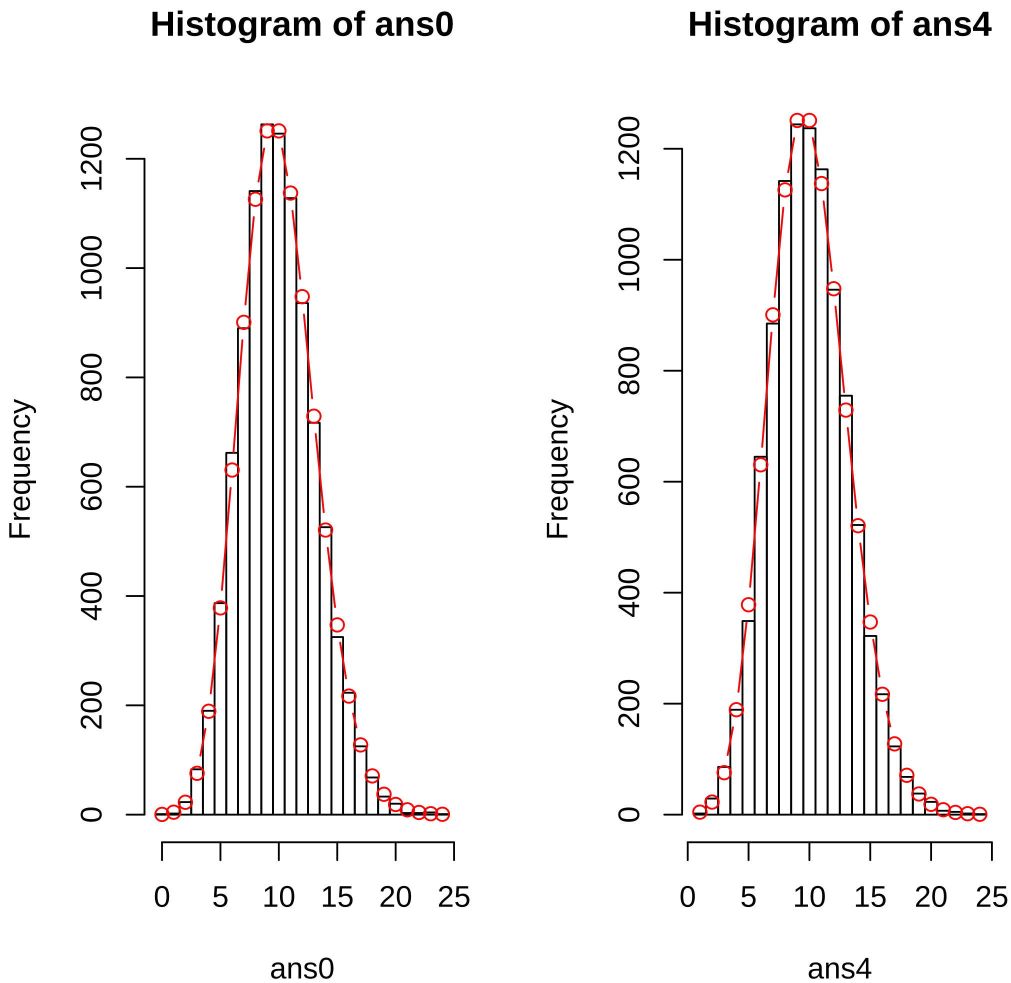
Solution: The R function `sapply` is a helpful way to obtain these answers. Since these are sample means, I'll report them as 95% confidence intervals:

```
> f0 <- function(vec) sum(0<vec & 1>vec)
> ans0 <- sapply(X, f0)
> mean(ans0) + c(-1, 1) * 1.96 * sd(ans0)/ sqrt(length(ans0))
[1]  9.891992 10.015608
> f4 <- function(vec) sum(4<vec & 5>vec)
> ans4 <- sapply(X, f4)
> mean(ans4) + c(-1, 1) * 1.96 * sd(ans4)/ sqrt(length(ans4))
[1]  9.936118 10.059882
```

(d) Plot a histogram each for the distribution of the number of events in the interval (0,1) and the interval (4,5) respectively, based on the 10,000 realizations.

Solution: Although it was not required, the code below produces histograms with the true theoretical Poisson counts superimposed:

```
> par(mfrow=c(1,2))
> hist(ans0, breaks=0.5+(min(ans0)-1):max(ans0))
> x0 <- min(ans0):max(ans0)
> lines(x0, 10000*dpois(x0, 10), col=2, type="b")
> hist(ans4, breaks=0.5+(min(ans4)-1):max(ans4))
> x4 <- min(ans4):max(ans4)
> lines(x4, 10000*dpois(x4, 10), col=2, type="b")
```



3. Cars pass a certain street location according to a Poisson process with rate λ. A woman who wants to cross the street at that location waits until she can see that no cars will come by in the next T time units.

   (a) Find the probability that her waiting time is 0.

   Solution: This is the probability that the first car will take longer than T to arrive, which is $e^{-\lambda T}$.

   (b) Find her expected waiting time.

   Solution: Let N be the number of cars that pass before she can cross. Since each passing car restarts the waiting, part (a) tells us that N is the same as the number of independent Bernoulli($e^{-\lambda T}$) trials before the first success occurs. We conclude that $N+1$ has a geometric distribution with mean $e^{\lambda T}$, so $E(N) = e^{\lambda T} - 1$.

   Now, for a car that passes in less than T time units, if X is the total waiting time, then $EX = E[E(X \mid N)]$. We know that $E(X \mid N)$ is simply N times the expectation of a single passing time, given that the time is less than T. To find the latter, we may integrate the conditional density of an exponential, say Y, given that $Y \le T$:

$$E(Y \mid Y \le T) = \frac{1}{1-e^{-\lambda T}} \int_0^T y\lambda e^{-\lambda y}\, dy = \frac{1-e^{-\lambda T}(\lambda T+1)}{\lambda(1-e^{-\lambda T})}$$

   We conclude that $E(X)$ is $E(N)$ times the above expression:

$$E(X) = \frac{(e^{\lambda T}-1)[1-e^{-\lambda T}(\lambda T+1)]}{\lambda(1-e^{-\lambda T})} = \frac{e^{\lambda T}-\lambda T-1}{\lambda}.$$