MATH 420 Homework #7 SOLUTIONS

1. (10 points) Look up a probability distribution that (1) doesn't appear on the 'Example Probability Distributions' handout circulated in class on Monday (see the course website for the PDF), and that (2) is frequently used in some scientific field of interest (e.g. geology, education, physics, psychology, chemistry, etc.). Feel free to use books, journal articles, "experts", etc.

Write up a brief (less than 1 page, but more than a few lines) summary of that distribution that (minimally) includes the same kinds of information as provided in each column of the handout. Include a 'References' section at the end that lists any references used in preparing your summary (format them appropriately).

2. (5 points) 7.3a (See pg.242 in the text.)

Using $a = 3 \cdot 10^{-9}$ and $T_n/n = 10^{-7}$ in the equation

$$\lambda = \frac{n}{T_n - n\,a} = \frac{1}{T_n/n - a}$$

yields

$$\lambda = \frac{1}{10^{-7} - 3 \cdot 10^{-9}} = 1.0309278 \cdot 10^{7}$$

3. (5 points) 7.3b

The Law of Large Numbers says that

$$\lim_{n\to\infty} \frac{T_n}{n} = a + \frac{1}{\lambda}$$

We interpret this to mean that, for sufficiently large n,

$$\frac{T_n}{n} \approx a + \frac{1}{\lambda}.$$

Similarly, the Central Limit Theorem implies that (for sufficiently large n)

$$\frac{T_n}{n} \sim \text{Normal}(\text{mean} = a + \frac{1}{\lambda},\ \text{std. dev.} = \frac{\sigma}{\sqrt{n}})$$

where $\sigma^2 = Var(X_k) = E((X_k - E(X_k))^2) = E((X_k - a - \frac{1}{\lambda})^2) = E((Y_k - \frac{1}{\lambda})^2) = Var(Y_k) = \frac{1}{\lambda^2}$.
Since 95% of the probability mass of a standard Normal distribution is between ± 1.959964, it follows that (for a true decay rate λ) the 95% range of normal variation is the interval

$$\left(a + \frac{1}{\lambda} - \frac{1.959964}{\sqrt{n}\,\lambda},\ a + \frac{1}{\lambda} + \frac{1.959964}{\sqrt{n}\,\lambda}\right)$$

4. (5 points) 7.4

For 7.3b, using a 99.7% level for observed values of T_n/n yields the interval

$$\left(a + \frac{1}{\lambda} - \frac{3}{\sqrt{n}\,\lambda},\ a + \frac{1}{\lambda} + \frac{3}{\sqrt{n}\,\lambda}\right)$$

Assuming that the textbook has a typo in part 7.3c, and that it should instead give the observed decay rate of $n/T_n = 10^7$ (note the units!) then over a period of 30sec we have $n = 30 \cdot 10^7 = 3 \cdot 10^8$ thus the range of λ values with 99.7% intervals that include this observed value lie between the λ values that satisfy

$$3 \cdot 10^{-9} + \frac{1}{\lambda} \pm \frac{3}{\sqrt{3 \cdot 10^8\,\lambda}} = 10^{-7} \quad \Leftrightarrow \quad \lambda = \frac{10^9 \pm \sqrt{3}\,10^5}{97}$$

That is, the range of λ values is centered at our estimate of λ from 7.3a (#2 above) and is approximately

$$(10307493,\ 10311064).$$

For 7.3d (at the 99.7% level) consider that we can rephrase the CLT result as

$$\frac{T_n}{n} - a \sim \text{Normal}\left(\text{mean } = \frac{1}{\lambda},\ \text{std. dev. } = \frac{1}{\sqrt{n}\,\lambda}\right)$$

and we further assume we can just invert this interval to answer the question. Then we have the interval for our estimate of $\frac{1}{\lambda}$ given by

$$\frac{1}{\lambda} \pm \frac{3}{\sqrt{n}\,\lambda} = \frac{\sqrt{n} \pm 3}{\sqrt{n}\,\lambda}$$

Inverting these two values gives the interval containing our estimate of λ as

$$\left(\frac{\lambda}{1 + 3/\sqrt{n}},\ \frac{\lambda}{1 - 3/\sqrt{n}}\right).$$

The threshold n values that would yield that estimates are within $0.5 \cdot 10^{-6}\lambda$ of λ are those n values larger than n satisfying

$$\frac{1}{1 - 3\sqrt{n}} - 1 = 0.5 \cdot 10^{-6}.$$

This yields $\sqrt{n} = 6000003$, or roughly a sample size of $3.6 \cdot 10^{13}$ decay events, or roughly $3.6 \cdot 10^6$ seconds, or 1000 hours! This fits the general rule of thumb that **you need to increase your sample size by two orders of magnitude for every decimal place of increased accuracy**. This "rule of thumb" isn't very sensitive to the given confidence level.

5. (5 points) 7.6b

Using R, we see that the probability of of calls received in a given month deviating from the mean of 171 by as much as 18 is approximately 84.3%.

```
# Using the CDF:
ppois(171 + 18, 171) - ppois(171 - 19, 171)

## [1] 0.8430889

# Alternatively, we could sum the probabilities directly:
sum(dpois((171 - 18):(171 + 18), 171))

## [1] 0.8430889
```

BONUS (3 points): 7.14a

Using the fact that $1 + x + x^2 + \cdots = (1-x)^{-1}$ for $x \in (0, 1)$, we have

$$P(X > i) = \sum_{k=0}^{\infty} p(1-p)^{i+k} = p(1-p)^i \sum_{k=0}^{\infty}(1-p)^k = p(1-p)^i p^{-1} = (1-p)^i$$

BONUS (3 points): 7.14b

By the definition of conditional probability (and 7.14a above)

$$P(X > i + j | X > j) = \frac{P(X > i + j)}{P(X > j)} = \frac{(1-p)^{i+j}}{(1-p)^j} = (1-p)^i = P(X > i)$$

BONUS (3 points): Compute the mean number of events in the interval [0,20] for 1000 replicates of the nonhomogeneous poisson process simulated in class on Monday (Nov 2). See course website for the R code from class.

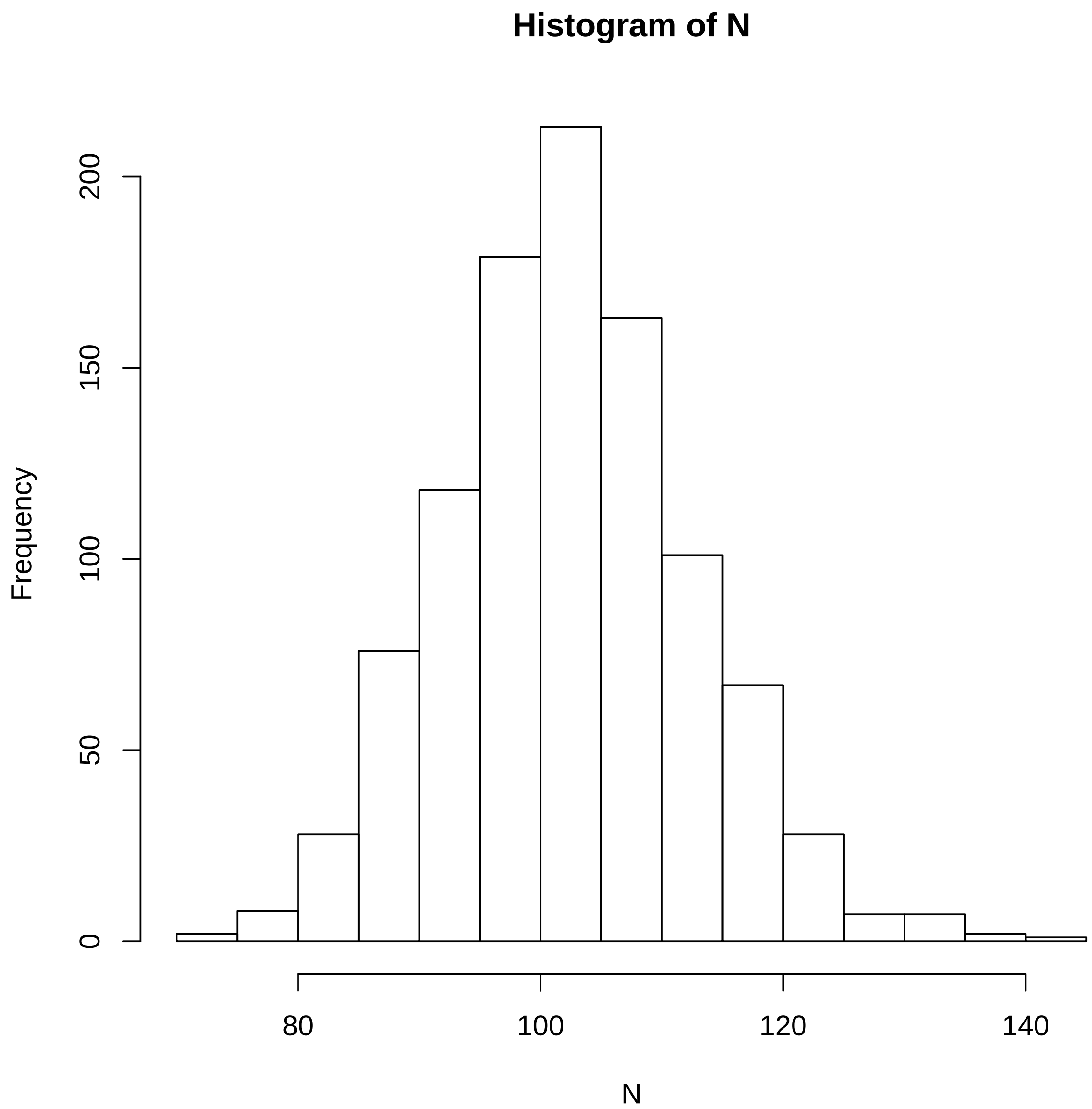
The theoretical result given in class states that

$$E(N_{[0,20]}) = \int_0^{20} r(t)\, dt.$$

Computing this expected value and simulating 1000 such realizations of the Inhomogeneous Poisson Process yields very similar results (as expected!):

```
## The code below simulates a homogenous Poisson Process, then thins the
## resulting set of events to simulate the given inhomogeneous P.P.  This is
## then replicated 1000 times and compared to the analytical expected value
## of the number of events in an interval (here, [0,20]).

## Set random number generator seed so we get identical results
set.seed(1)
```

```r
## Inhomogeneous Poisson Process definition First the rate function r(t) = r0
## + A sin(w t)
r0 = 5
A = 4.9
w = 1
rmax = r0 + A
r = function(t) {
    r0 + A * sin(w * t)
}

## 1000 replicate simulations of the IPP to get an average count of events on
## the interval [0,20]. Store those counts in N
N = c()
for (j in 1:1000) {
    # Simulate at rmax=r0+A=9.9 and thin as we go until we reach time=20.
    Tk = c(rexp(1, rate = rmax))
    Sk = cumsum(Tk)
    while (tail(Sk, 1) < 20) {
        Tk = c(Tk, rexp(1, rate = rmax))
        Sk = cumsum(Tk)
    }

    ## Remove the last event, which occurs after time=20
    Tk = head(Tk, length(Tk) - 1)
    Sk = cumsum(Tk)

    # Thin the process using the vector-based approach. For more on thinning,
    # see www.pauljhurtado.com/teaching/FA15/poisson-process.R

    # Vector of P values to determine whether to keep each event (or not) Note
    # Ps=r(Sk) is the same as for(i in 1:length(Sk)) { Ps[i]=r(Sk[i]) }
    Ps = r(Sk)/(r0 + A)

    # Vector of random uniform [0,1] (see for loop in 1st approach above)
    RNs = runif(length(Sk))

    # Determine whether to keep each even (or not)
    keep = (RNs <= Ps)  # parentheses not needed. Returns vector of T/F. TRUE=KEEP.

    # Keep the appropriate events
    Skthinned = Sk[keep]

    # Compute the number of events in [0,20]
    N[j] = length(Skthinned)  # N[j] = sum(keep) also works!

}  ## End for() loop

## Histogram and compare E(N) (integral) vs. simulated mean
hist(N)
```




```
data.frame(Expected = mean(N), Simulated = integrate(r, 0, 20)$value)
```

```
##   Expected Simulated
## 1  102.869  102.9004
```

```
## The LLN assures us these numbers should be close!
```