Stat 135, Fall 2006 A. Adhikari

HOMEWORK 3 SOLUTIONS

1. From Example A you know that the distribution of $\lambda_0 - \hat{\lambda}$ is approximately normal with mean 0 and sd 1.04. The chance in the problem is equal to two equal tails of that normal histogram, outside the range $\pm\delta$. The approximate chances are, respectively: 0.63, 0.34, 0.15, 0.055, 0.0164.

2. The MOM estimate of λ is the sample mean:

$$\frac{0 \times 14 \ + \ 1 \times 30 \ + \ 2 \times 36 \ + \cdots + \ 1 \times 12}{300} \ = \ 3.893$$

Plug this value into the Poisson probability formula to get the estimated proportions in each category; then multiply by 300 to get the estimated counts. I used R:

```
> dpois(0:12, lambda=3.893)*300
```

gives estimated counts of 6.115, 23.807, 46.340, 60.133, 58.525, 45.567, 29.566, 16.443, 8.001, 3.461, 1.347, 0.477, and 0.155.

The fit is clearly underestimating the number of intervals with low counts. The Poisson model assumes that the rate of arrivals is constant over time, whereas of course it's not constant in reality. The rate of arrivals at night, and possibly on the weekends depending on where the intersection is, is likely to be lower than at other times. This will result in a greater observed number of low-count intervals.

3a. The first population moment is $\mu_1 \ = \ \theta/3 \ + \ 4(1-\theta)/3 \ + \ (1-\theta) \ = \ 7/3 - 2\theta$. So the MOM estimate of θ is

$$\hat{\theta} \ = \ \frac{7}{6} - \frac{\hat{\mu}_1}{2} \ = \ \frac{7}{6} - \frac{\bar{X}}{2}$$

In our sample the observed value of $\bar{X}$ is 1.5, so the estimate is $5/12 = 0.4167$.

b. $Var(\hat{\theta}) \ = \ \frac{1}{4}Var(\bar{X}) \ = \ \frac{1}{4}\frac{\sigma^2}{10}$ where σ^2 is the population variance. The 10 is the sample size. Now all we need is σ^2. For this:

$$\mu_2 \ = \ \frac{1}{3}\theta \ + \ \frac{8}{3}(1-\theta) \ + \ \frac{9}{3}(1-\theta)$$

and we already have μ_1. Now $\sigma^2 = \mu_2 - \mu_1^2$. This is a function of θ so to approximate it you can plug in the estimate 0.4167 for θ. The approximate variance is 0.03 and the approximate SE is 0.17.

You could also estimate σ^2 by $\hat{\sigma}^2$ from the sample. With that method the approximate SE comes out to be 0.16.

4a.

$$\begin{aligned}
\Gamma(\alpha+1) &= \int_0^\infty t^\alpha e^{-t} dt \\
&= t^\alpha(-e^{-t})|_{t=0}^\infty \ - \ \int_0^\infty \alpha t^{\alpha-1}(-e^{-t})dt \\
&= \alpha \int_0^\infty t^{\alpha-1}e^{-t}dt \ = \ \alpha\Gamma(\alpha)
\end{aligned}$$

b. Let $Y = Z^2$, and let ϕ be the standard normal density. Then by the change of variable formula for densities (see e.g. Pitman Example 5 page 307) the density f_Y of Y is obtained by

$$f_Y(y) \ = \ [\phi(\sqrt{y}) + \phi(-\sqrt{y})]/2\sqrt{y} \ = \ \frac{2}{\sqrt{2\pi}}e^{-\frac{1}{2}y}/2\sqrt{y} \ = \ \frac{1}{\sqrt{2\pi}}y^{\frac{1}{2}-1}e^{-\frac{1}{2}y}$$

You know that the change of variable formula gives you a density, i.e., a non-negative function that integrates to 1. Look at the functional form (the bits that involve y), ignoring the constant. That's the form of the gamma density with parameters $\alpha = 1/2$ and $\lambda = 1/2$. So the constant $1/\sqrt{2\pi}$ must agree with the constant in its general form:

$$\frac{1}{\sqrt{2\pi}} = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})}$$

So $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

This problem can also be solved by differentiating the normal c.d.f. This is worked out, with all bells and whistles, in Example C on page 61 of Rice.

5. The non-existent problem. Enjoy.

The solutions below consist of suggested R code that will work. These methods are easy to understand, but I expect many of you to find methods that are swifter and better. On this assignment you were not asked to turn in code, so I expect your answers to consist just of the final results. If you turned in your code as well, that's fine. Answers may be handwritten or printed out, but plots must be printouts of R output.

6. I pretended that my SID number is 12345678. Hence:

```
> pop < − c(1, 2, 3, 4, 5, 6, 7, 8)
> mean(pop)      # population mean = 4.5
> sd(pop)*sqrt(7/8)      # population SD = 2.291288
> hist(pop, breaks=seq(0.5, 8.5, by=1, prob=TRUE)
# center the bars over the values, and plot the proportions not the counts
```

7. No, you don't have to draw a sample at all. The expectation and variance are both familiar results. The expected value of the sample mean is the population mean, which is 4.5. The standard error of the sample mean is $\sigma/\sqrt{400}$, where σ is the population SD which is 2.291288. So the standard error of the sample mean is 0.1145644. And the distribution of the sample mean is roughly normal by the CLT.

To plot this normal curve, first note that the values of the mean and SE imply that the action is almost all going to be over the 4 to 5 range.

```
> x ← seq(4, 5, by=0.1)      # values at which to calculate the density
> plot(x, dnorm(x, mean=4.5, sd=0.1145644), type=”l”)
```

8. First generate 100 samples of size 400 each:

```
> bigsample ← sample(pop, size=40000, replace=TRUE)
> samples ← matrix(bigsample, nrow=400, ncol=100)
```

Each column of *samples* is a sample of size 400. So now

```
> sampmeans ← colMeans(samples)
> hist(samples, prob=TRUE)
> lines(x, dnorm(x, mean=4.5, sd=0.1145644))
```

The old x from Exercise 7 should work, because you're drawing exactly the same normal density. But I know that many of you will change the x-values a bit, depending on the max and min of your sample means.

9. Get each sample SD and convert it to the estimated SE for the sample mean:
> sampses ← apply(samples, 2, sd)*sqrt(299/300)/sqrt(400)
(Yes, it's fine if you didn't use the $sqrt$299/300.) Now get the two ends of the intervals, and combine to get one 100x2 matrix of intervals:
> leftends ← sampmeans - 1.96*sampses
> rightends ← sampmeans + 1.96*sampses
> intervals ← cbind(leftends, rightends)
Now select the bad intervals, that is, the ones that don't contain the population mean.
> bad ← intervals[intervals[,1] > 4.5 — intervals[,2] < 4.5,]
> bad
That last command will show you the bad intervals. The number of good intervals you got is 100 - number bad.

10. Each confidence interval contains the parameter with chance 0.95, independently of the other intervals. So the distribution of G is binomial (100, 0.95). Its expectation is 95 and its standard error is $\sqrt{100 \times 0.95 \times 0.05} = 2.18$. So essentially all the probability will be over the 80-100 range. Therefore, using a clunky way (outlined on the web) to get a histogram from a table of probabilities:
> x ← 80:100
> probs ← dbinom(x, size=100, p=0.95)
> xplus ← c(x-.5, 100.5)
> probsplus ← c(probs, probs[21])
> plot(xplus, probsplus, type="s")
> lines(xplus, probsplus, type="h")
If you got 93 good intervals:
> points(93, 0, pch="25")
will give you a triangle in the right place.

There are 80 students in the class. So it's not unlikely that someone will get an unusually bad set of samples and therefore an unusually small number of good intervals. Let's see!