STA 200A: Homework 9; Solution

This assignment will not be collected for credit.

1. 7.P50

Solution: Notice that $f(x, y)$ can be expressed in the form $C_y e^{-x/y}$ where C_y is a function of y alone. So, using the fact that if two pdfs are proportional then they are equal, we conclude that the conditional distribution of X given $Y = y$ has an Exponential(λ) distribution, where $\lambda = 1/y$. This has second moment equal to $2/\lambda^2 = 2y^2$.

2. 7.T50

Solution:

$$\frac{\partial}{\partial t}\log M_X(t) = \frac{M_X'(t)}{M_X(t)}, \quad \frac{\partial^2}{\partial t^2}\log M_X(t) = \frac{M''(t)}{M(t)} - \frac{(M'(t))^2}{(M(t))^2}.$$

Notice that $M_X(0) = 1$, $M_X'(0) = E[X]$, $M_X''(0) = E[X^2]$, so

$$\Psi''(0) = E[X^2] - (E[X])^2 = Var(X).$$

3. 8.P4

Solution: Call $R = \sum_{i=1}^{20} X_i$ then $E[R] = 20$ and $Var(R) = 20$.

(a)

$$P\{R > 15\} < \frac{20}{15} = \frac{4}{3}.$$

This is not very helpful, because it merely says that the probability is less than a number bigger than 1.

(b)

$$P\{R > 15\} = P\{\frac{R-20}{\sqrt{20}} > \frac{15-20}{\sqrt{20}}\} \approx P\{Z > -1.12\} = 0.87.$$

4. 8.P9

Solution: We could do this in a number of ways. Let's try Chebyshev's inequality,

$$P\{|X/n - 1| > .01\} < Var(X/n)/.01^2 = 10000/n,$$

which is at most .01 when $n \geq 1,000,000$. At this scale the normal approximation is reasonable to use, so let's try that, we want

$$P\{|(X-n)/\sqrt{n}| > .01\sqrt{n}\} \approx .01$$

by the 0.99 quantile of the standard normal is about 2.33 i.e. $P(Z \geq 2.33) \approx 0.01$, so we need that

$$\sqrt{n} \geq 2.33/.01 = 233.$$

i.e. $n \geq (233)^2 = 54,289$. This is significantly better than what we got from Chebyshev.

5. 8.P14

Solution: Suppose that we start with N components with lifetimes T_i. Then we want that

$$P\{\sum_{i=1}^{N} T_i > 2000\} \geq 0.95$$

but this requires that

$$P\{\sum_{i=1}^{N} \frac{T_i - 100}{30\sqrt{N}} > \frac{2000 - 100N}{30\sqrt{N}}\} \geq 0.95.$$

Let $Z \sim N(0,1)$. It is a basic fact that .05 quantile for Z is approximately -1.645. In other words $P(Z \geq -1.645) \approx 0.95$. By the CLT, we have that $\sum_{i=1}^{N} \frac{T_i - 100}{30\sqrt{N}}$ is approximately distributed according to Z. Hence, as long as

$$\frac{2000 - 100N}{30\sqrt{N}} \leq -1.645. \quad (*)$$

then the CLT indicates that we will have $P\{\sum_{i=1}^{N} T_i > 2000\} \geq 0.95$ (up to some approximation error). Define $S = \sqrt{N}$ then (*) implies

$$\frac{200}{3} - \frac{10}{3}S^2 \leq -1.645S$$

i.e.

$$0 \leq \frac{10}{3}S^2 - 1.645S - \frac{200}{3}.$$

Let's solve the quadratic equation

$$3.334S^2 - 1.645S - 66.67 = 0$$

which leads to

$$S \geq 4.72.$$

which is roughly the same as $N \geq 22$.

6. 8.P19

Solution: Consider the variables $Y_i, i = 1, \ldots, 4$ where Y_i is the time since the last observed new type until the next observed new type. Furthermore, $Y_1 = 1$ and each other Y_i are independent Geometric(p_i) random variables with $p_2 = 3/4, p_3 = 1/2, p_4 = 1/4$.

(a) Let $Y = Y_1 + Y_2 + Y_3 + Y_4$ then,

$$E[Y] = 1 + \frac{4}{3} + 2 + 4 = 8.34, \quad Var(Y) = 0 + \frac{1-\frac{3}{4}}{(3/4)^2} + \frac{1-\frac{1}{2}}{(1/2)^2}\frac{1-\frac{1}{4}}{(1/4)^2} = 14.45.$$

So we can use Chebyshev's inequality to show that

$$P\{|Y - 8.34| > t\} \leq \frac{14.45}{t^2}.$$

Setting this equal to .1 gives us $t = \sqrt{14.45/0.1} = 12.02$, So $-3.68 = 8.34 - 12.02 \leq Y \leq 8.34 + 12.02 = 20.36$ with probability at least 0.9. But because $Y > 1$ almost surely, we can use $a = 1, b = 20.36$.

(b) We have that

$$P\{Y \geq a\} \leq \frac{14.45}{14.45 + a^2} = .1$$

gives $a = 11.4$. This is better than what Chebyshev's gives us in (a).

7. 8.P23

Solution:

(a) $p \leq 20/26 = 0.77$.

(b) $p \leq \frac{20}{20+6^2} = 0.36$.

(c) $p \leq e^{-20}(20e)^{26}/26^{26}$. This is going to be difficult to compute due to machine tolerance, so let's take a log! $\log p \leq -20 + 26\log(20e) - 26\log 26 = -.821$ so $p \leq 0.44$.

(d) The first question to ask is, why would the CLT work in this case? Are we summing iid random variables? Why yes, if Y_i are independent Poisson(1), then $Y = \sum_{i=1}^{20} Y_i$. Then the CLT says that

$$P\{(Y-20)/\sqrt{20} > (26-20)/\sqrt{20}\} \approx P\{Z > 1.34\} = 0.09.$$

(e) Here is some R code...

```
> T = 10000
> X = rpois(T,20)
> mean(X >= 26)
[1] 0.1106
```

So we have a simulated probability of 0.11.

8. 8.T8

Solution: When t is an integer, note thatGamma(t, λ) is the sum of t iid exponential(λ) random variables.

9. A fair coin is flipped repeatedly until 50 heads are observed. What is the probability that at least 80 flips are necessary? (You may calculate an approximate answer.)

Solution: Recall that a negative binomial distribution, represents the number of attempts to reach k successes, where each attempt has success probability p. In this case, let X be a negative binomial variable based on $k = 50$ and $p = 1/2$ (fair coin).

A useful fact to note about the negative binomial distribution is that can be represented in the following way. Let $Y_1, \ldots, Y_k$ be independent Geometric(p) random variables. Then, it can be seen that

$$X = Y_1 + Y_2 + \cdots + Y_k.$$

The reason, is that each Y_2 represents the number of "extra" attempts beyond the first success, and Y_3 represents the number of "extra" attempts beyond the second success, and so on.

By the central limit theorem, we have that X approximately follows the distribution $N(\mu, \sigma^2)$, where

$$\mu = E[X] = E[Y_1 + \cdots + Y_k] = kE[Y_1] = \frac{k}{p},$$

(Recall that $E[Y_1] = 1/p$ for a Geometric(p) variable.), and

$$\sigma^2 = var(X) = var(Y_1 + \cdots + Y_k) = kvar(Y_1) = k\frac{1-p}{p^2}.$$

Therefore, plugging in $k = 50$ and $p = 1/2$, we get $\mu = 100$, and $\sigma^2 = 100$.

Hence, if let $Z \sim N(0, 1)$, then $\sqrt{100}Z + 100 \sim N(\mu, \sigma^)$, and we get from the CLT that

$$P(X \geq 80) \approx P(\sqrt{100}Z + 100 \geq 80) = P(Z \geq \frac{80 - 100}{\sqrt{100}}) = 1 - \Phi(-2) \approx 0.98.$$