Introduction to Probability

Math 30530, Section 01 — Fall 2012

Homework 9 — Solutions

1. During a season, a basketball team plays 70 home games and 60 away games. The coach estimates at the beginning of the season that the team will win each home game with probability .7, and each away game with probability .4, all games independent. *Estimate* the probability that, under these assumptions, the team wins at least 80 games in total during the season. (Don't try to calculate the exact probability; that's quite a pain).

   Solution: Let X be number of home games won, Y number of road games won. Have $X \sim$ Binomial(70, .7) and $Y \sim$ Binomial(60, .4). We can't directly add X and Y because the p's are different; but we can approximate using the DeMoivre-Laplace theorem and then add the resulting approximations. We have $X \approx \mathcal{N}(49, 14.7)$ and $Y \approx \mathcal{N}(24, 14.4)$, so $X + Y \approx \mathcal{N}(73, 29.1)$. We want to estimate the probability that $X + Y \geq 80$. Since we are approximating an integer valued random variable with a normal, we should use the continuity correction and calculate

   $$\Pr(\mathcal{N}(73, 29.1) > 79.5) = .1141\ldots.$$

2. Melons at the farmers market have a weight that is normally distributed with mean 3lbs and standard deviation .4lbs. The plastic bags that the melon vendor uses have breaking weight that is normally distributed with mean 10lbs an standard deviation .6lbs. Calculate the probability that is three melons are put into a bag, the bag will break (i.e., the breaking weight will be exceeded).

   Solution: If X_i is the weight of the ith melon, and Y the breaking weight of the bag, we want

   $$\Pr(X_1 + X_2 + X_3 > Y) = \Pr(X_1 + X_2 + X_3 - Y > 0).$$

   We have $X_1+X_2+X_3-Y \sim \mathcal{N}(3+3+3-10, (.4)^2+(.4)^2+(.4)^2+(.6)^2) = \mathcal{N}(-1, .84)$ (notice that we *add* the variance of Y, because $(-1)^2 = 1$). We want

   $$\Pr(\mathcal{N}(-1, .84) > 0) = .1376\ldots.$$

3. I keep flipping a coin until I have seen heads 200 times. *Estimate* the probability that this process will take me between 380 and 420 flips, inclusive (for this problem, you will want to use the continuity correction).

Solution: Let X be number of flips; $X \sim$ NegBinomial$(200, .5)$. To estimate this note that it is the sum of 200 independent geoemetrics each with $p = .5$, so by central limit theorem $X \approx \mathcal{N}(200 \times (1/.5), 200 \times (1-.5)/(.5^2)) = \mathcal{N}(400, 400)$. Using the continuity correction, we estimate

$$\Pr(380 \leq X \leq 420) \approx \Pr(379.5 \leq \mathcal{N}(400,400) \leq 420.5) = .84732\ldots - .15268\ldots = .69464\ldots.$$

4. (a) Verify that if X and Y are independent exponential random variables, each with parameter λ, then $Z = X + Y$ has density function

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z} & \text{if } z \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Solution: We did this in class, and it is covered as a special case of the next part, so I won't write the solution here.

(b) Suppose that X is an exponential random variable with parameter λ, and Y is a Gamma random variable with parameters r and λ, and that X and Y are independent. Verify that $Z = X + Y$ has density function

$$f_Z(z) = \begin{cases} \frac{\lambda^{r+1}}{r!} z^r e^{-\lambda z} & \text{if } z \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

(This proves, by induction, that the formula for the Gamma density is correct.)

Solution: The joint density of X and Y is

$$\lambda e^{-\lambda x} \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y}$$

on the first quadrant, and 0 elsewhere. The range of possible values for $X + Y$ is 0 to ∞, so both the CDF and density of $X + Y$ are 0 is the argument z is less than 0. If $z > 0$ then

$$P(X + Y \leq z) = \int_0^z \int_0^{z-x} \lambda e^{-\lambda x} \frac{\lambda^r}{(r-1)!} y^{r-1} e^{-\lambda y} \, dy dx.$$

This is hideous, because the inner integration involves $y^{r-1} e^{-\lambda y}$, so we try switching the order of integration, to at least be able to do the inner integration.

We have

$$\begin{aligned}
P(X+Y\le z) &= \int_0^z\int_0^{z-y}\lambda e^{-\lambda x}\frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dxdy\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\left[-e^{-\lambda x}\right]_0^{z-y}\,dy\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\left(1-e^{-\lambda z+\lambda y}\right)\,dy\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dy-\int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda z}\,dy\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dy-e^{-\lambda z}\int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}\,dy\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dy-e^{-\lambda z}\left[\frac{\lambda^r}{r!}y^r\right]_0^z\\
&= \int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dy-\frac{\lambda^r}{r!}z^re^{-\lambda z}.
\end{aligned}$$

Normally we would at this point complete the integration to find the CDF, then differentiate to find the density; but the integration doesn't seem too nice here. But there' a fix: we can use fundamental theorem of calculus to differentiate with respect to z, without first calculating the integral! Specifically,

$$\frac{d}{dz}\int_0^z \frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}\,dy=\frac{\lambda^r}{(r-1)!}z^{r-1}e^{-\lambda z}$$

by FTOC. Also, by product rule,

$$\frac{d}{dz}\left(\frac{\lambda^r}{r!}z^re^{-\lambda z}\right)=-\frac{\lambda^{r+1}}{r!}z^re^{-\lambda z}+\frac{\lambda^r}{(r-1)!}z^{r-1}e^{-\lambda z}.$$

Combining these we get that for $z\ge 0$,

$$\frac{d}{dz}\Pr(X+Y\le z)=f_{X+Y}(z)=\frac{\lambda^{r+1}}{r!}z^re^{-\lambda z},$$

exactly as required.

5. **GW** 33.2 (not part g))

   a) We want to know the time until the rth occurrence of an event, when occurrences are independent, simultaneous occurrences are very unlikely, and the average number of occurrences per unit time is known - this is exactly to setup for the Gamma random variable.

   b) X is the time (in years) until the third occurrence of category 4 or higher hurricane.

   c) $r=3$, $\lambda=1/6$ (one every 6 years).

d) $E(X) = r/\lambda = 18$.

e) $\mathrm{Var}(X) = r/\lambda^2 = 108$.

f) For $x < 0$, $f(x) = 0$. For $x \geq 0$,

$$f(x) = \frac{(1/6)^3}{2!} x^2 e^{-x/6}.$$

g)

$$\Pr(X \leq 10) = \int_0^{10} \frac{(1/6)^3}{2!} x^2 e^{-x/6} \, dx = .234\ldots.$$

6. **GW** 33.11

   Modeling the number of minutes as a Gamma random variable with $r = 500$, $\lambda = 10$ (why 10? because the density given for a single gap is exactly an exponential with $\lambda = 10$), we get that the expectation is $500/10 = 50$ minutes.

7. **GW** 33.15

   We model the sum of the three waiting times as the sum if three independent exponentials, each with parameter $1/10$ (so average 10); i.e. as a Gamma with $r = 3$, $\lambda = 1/10$. The probability we want is then

$$\int_{40}^{\infty} \frac{(1/10)^3}{2!} x^2 e^{-x/10} \, dx = .2381\ldots.$$

8. **GW** 36.1

   Let X be the total time; X is the sum of four independent normals, each with mean 23.8 and variance 25, so X is normal with mean $4 \times 23.8 = 95.2$ and variance $4 \times 25 = 100$. So the probability we want is

$$\Pr(\mathcal{N}(95.2, 100) < 90) = .30153\ldots.$$

9. **GW** 36.10

   a) Here we model the total length as $X \sim \mathcal{N}(10000, 20 \times (75)^2)$ and we want

$$\Pr(\mathcal{N}(10000, 112500) < 10560) = 9525\ldots.$$

   b) The total length is now $X \sim \mathcal{N}(10000 + (13.5)(21), 20 \times (75)^2)$ and so we want

$$\Pr(\mathcal{N}(10283.5, 112500) < 10560) = .80101\ldots.$$

   c) Here the total length of the gaps is $X \sim \mathcal{N}(10000, 20 \times (75)^2)$, and the total length of the cars is $Y \sim \mathcal{N}(20(13.5), 20 \times (1)^2)$, and so the total length is $\mathcal{N}(10000 + 20(13.5), 20 \times (75)^2 + 20 \times (1)^2)$. We want

$$\Pr(\mathcal{N}(10283.5, 112520) < 10560) = .80099\ldots.$$

10. **GW** 36.13

    The average of the 20 jumps is a normal random variable with mean 7 and variance $(20 \times (.2)^2)/(20^2) = .002$. So we want

    $$\Pr(6.95 \leq \mathcal{N}(7, .002) \leq 7.05) \approx 1.$$

11. **GW** 37.3

    If X_i is his RBI count for each game, then X_i has mean .7, variance .04. If X is his number of RBIs for the season, then X has mean $162 \times .7 = 113.4$ and variance $162 \times .04 = 6.48$. By CLT, $X \approx \mathcal{N}(113.4, 6.48)$. Since RBIs come in whole numbers only, we must use a continuity correction. The probability we want is

    $$\Pr(X \leq 110) \approx \Pr(\mathcal{N}(113.4, 6.48) \leq 110.5) = .8726\ldots.$$

12. **GW** 37.10 (the length of the shower is a random time)

    BY CLT, time spent in shower over 45 days is approximately $\mathcal{N}(45 \times 15, 45 \times (4^2)) = \mathcal{N}(675, 720)$, and the probability we want is

    $$\Pr(660 \leq \mathcal{N}(675, 720) \leq 720) = .66519\ldots.$$

    (Note no continuity correction here, since the original random variable is continuous.)

13. **GW** 37.34

    If X is the exactly number of requests for pop songs, then $X \sim \text{Binomial}(120, .37)$. By the DeMoivre-Laplace theorem, $X \approx \mathcal{N}(44.4, 27.972)$. Recall the continuity correction, we want

    $$\Pr(X \geq 50) \approx \Pr(\mathcal{N}(44.4, 27.972) \geq 49.5) = .1675\ldots.$$

14. **GW** 37.38 (for both parts, what is sought is an approximate answer)

    For both parts, let X be number of students attending on a given day; $X \sim \text{Binomial}(1500, .3) \approx \mathcal{N}(450, 315)$.

    a)

    $$\Pr(X > 500) \approx \Pr(\mathcal{N}(450, 315) > 500.5) = 99788.$$

    b) Suppose the lecture hall has n seats. The probability of overflow is

    $$\Pr(X > n) \approx \Pr(\mathcal{N}(450, 315) > n + .5) = \Pr\left(Z > \frac{n + .5 - 450}{\sqrt{315}}\right).$$

    We want this to be at most .05, which means we want

    $$\frac{n + .5 - 450}{\sqrt{315}} \geq 1.645$$

    or $n \geq 478.69\ldots$. Since n must be an integer, the smallest possible value of n is $n = 479$.

15. **GW** 37.49 (this question requires using a continuity correction)

    The number of mistakes he makes on 50 homeworks, on average, is 100; so a good model for the total number of mistakes is a Poisson with $\lambda = 100$. We approximate this by a normal with mean and variance 100. Recall the continuity correction (always, the number of mistakes made is a whole number), what we are looking for to approximate the probability of strictly more than 105 mistakes is

$$\Pr(\mathcal{N}(100, 100) \geq 105.5) = .2911\ldots.$$