Assignment 7

15-359 Probability and Computing
Name: Shashank Singh
Email: sss1@andrew.cmu.edu
Section: B
Due: Friday, March 30, 2012

Problem 1: Spin me right around

A. Since X and Y are independent,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \cdot \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} = \boxed{\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}.}$$

B. If r is the distance of (x, y) from the origin, $r = x^2 + y^2$, so that

$$f_{X,Y}(x,y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}} = \frac{1}{2\pi}e^{-\frac{r^2}{2}},$$

so that $f_{X,Y}(x,y)$ depends only on r. ■

Problem 2: Unwieldy numbers

Number the tweets $1, 2, \ldots, 340 \times 10^6$ in some order. $\forall i \in \{1, 2, \ldots, 340 \times 10^6\}$, let L_i be a random variable denoting the length of the i^{th} tweet. Since $L_i \sim \text{Uniform}(10, 140)$, $E[L_i] = \frac{10+140}{2} = 75$, and $\text{Var}[L_i] = \frac{(140-10)^2}{12} = 4225/3$. Let L be a random variable denoting the total number of characters processed by Twitter on an average day, so that $L = \sum_{i=1}^{340\times10^6} L_i$. Then, since 340×10^6 is large, and $L_1, L_2, \ldots, L_{340\times10^6}$ are independent and identically distributed, by the Central Limit Theorem, we can approximate the distribution of $\frac{L-n\mu}{\sigma\sqrt{n}}$ with $\mathcal{N}(0, 1)$, and therefore we can approximate the distribution of L as

$$L \sim \mathcal{N}(n\mu, \sigma^2 n) = \mathcal{N}(340 \times 10^6 * 75, 340 \times 10^6 * 4225/3).$$

Integrating the probability density function numerically from 25×10^9 to 26×10^9 gives

$$P(25 \times 10^9 \leq L \leq 26 \times 10^9) \approx 1.$$

Problem 3: Exp farming

As shown in class, $\widetilde{X}(s) = \frac{\lambda}{s+\lambda}$. It can be shown by induction on k that, $\forall k \in \mathbb{N}$,

$$\frac{d^k}{ds^k}\left(\frac{\lambda}{s+\lambda}\right) = (-1)^k \frac{k!\lambda}{(s+\lambda)^{k+1}}.$$

Thus, the k^{th} moment of X is given by

$$\begin{aligned} E[X^k] &= (-1)^k \frac{d^k\,\widetilde{X}(s)}{d\,s^k}\bigg|_{s=0} \\ &= (-1)^{2k} \frac{k!\lambda}{(s+\lambda)^{k+1}}\bigg|_{s=0} \\ &= \frac{k!\lambda}{\lambda^{k+1}} = \boxed{\frac{k!}{\lambda^k}.} \end{aligned}$$

Problem 4: Random amounts of randomness

1. Since $X_1, X_2, \ldots, X_{10}$ are independent, $e^{-sX_1}, e^{-sX_2}, \ldots, e^{-sX_{10}}$ are independent as well, and, since $X_1, X_2, \ldots, X_{10}$ are identicaly distributed, $\widetilde{X}_1 = \widetilde{X}_2 = \ldots = \widetilde{X}_{10} = \widetilde{X}$. Thus,

$$\widetilde{S}(s) = E\left[e^{-sS}\right] = E\left[e^{-s\left(\sum_{i=1}^{10} X_i\right)}\right] = E\left[\prod_{i=1}^{10} e^{-sX_i}\right] = \prod_{i=1}^{10} E\left[e^{-sX_i}\right] = \prod_{i=1}^{10} \widetilde{X}_i(s) = \boxed{\left(\widetilde{X}(s)\right)^{10}.}$$

2. By definition of the z-transform and the Laplace Transform,

$$\widehat{N}\left(\widetilde{X}(s)\right) = E\left[\left(\widetilde{X}(s)\right)^N\right] = E\left[E\left[e^{-sX}\right]^N\right].$$

Since $X_1, X_2, \ldots, X_i$ are i.i.d., $e^{-sX_1}, \ldots, e^{-sX_n}$ are i.i.d. as well, so that

$$\widehat{N}\left(\widetilde{X}(s)\right) = E\left[E\left[e^{-sXN}\right]\right] = E\left[e^{-s\sum_{i=1}^{N} X_i}\right] = E\left[e^{-sS}\right] = \widetilde{S}(s) \quad \blacksquare.$$

3. By the result of part 2. above, $\widetilde{S}(s) = \widehat{N}(\widetilde{X}(s))$. Thus, since we know the z-transform of Geometric(p) and the Laplace Transform of Exp(μ),

$$\widetilde{S}(s) = \widehat{N}\left(\widetilde{X}(s)\right) = \frac{\widetilde{X}(s)p}{1-\widetilde{X}(s)p} = \frac{\left(\frac{\mu}{\mu+s}\right)p}{1-\left(\frac{\mu}{\mu+s}\right)p} = \frac{\mu p}{\mu+z-\mu(1-p)} = \frac{\mu p}{s+\mu p}.$$

Thus, the Laplace Transform of S is that of Exp(μp), so that $S \sim$ Exp(μp). $\blacksquare$

Problem 5: Mouse in a maze

1. Let X, Y, and Z be random variables, with X denoting the time the mouse takes to return to his starting position if he goes left (so that $X \sim \text{Exp}(1)$), Y denoting the time the mouse takes if he goes right and then leaves (so that $Y \sim \text{Exp}(2)$), and Z denoting the time the mouse takes if he goes right and then returns to his starting position (so that $Z \sim \text{Exp}(3)$). Then,

$$E[T] = \frac{1}{2}(X + E[T])) + \frac{1}{2}\left(\frac{1}{3}Y + \frac{2}{3}(Z + E[T])\right).$$

Furthermore,

$$E[T] = E[E[T]] = E\left[\frac{1}{2}(X + E[T])) + \frac{1}{2}\left(\frac{1}{3}Y + \frac{2}{3}(Z + E[T])\right)\right] = \frac{25}{36} + \frac{5}{6}E[T],$$

so that, solving for $E[T]$ gives $E[T] = \boxed{\frac{25}{6}.}$

Problem 6: Chernoff proof relents

Let X_p and X_q be random variables, counting the number of heads among n coin flips each of two coins, with probabilities p and q or getting heads, respectively (where $p > q$). Then, since $q > p$ (and, consequently, $1 - q < 1 - p$), $\forall k \geq nq$,

$$\begin{aligned}
\frac{P(X_p = k)}{P(X_q = k)} &= \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k}q^k(1-q)^{n-k}} \\
&= \left(\frac{p}{q}\right)^k \left(\frac{1-p}{1-q}\right)^{n-k} \\
&= \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{n-qn} \\
&\leq \left(\frac{p}{q}\right)^{qn} \left(\frac{1-p}{1-q}\right)^{n-qn} \\
&= e^{-n\text{KL}(q,p)}
\end{aligned}$$

Thus, taking the sum over all $k \geq qn$ gives

$$P(X_p \geq qn) \leq P(X_q \geq qn)e^{-n\text{KL}(q,p)} \leq e^{-n\text{KL}(q,p)},$$

since $P(X_q \geq qn)$ is a probability and therefore $P(X_q \geq qn) \leq 1$, proving the desired result. ■

Problem 7: Polling is cheap. (Tell that to a systems programmer.)

Let p be the actual fraction of people who disapprove of the president. Then, assuming it is possible for people to be polled multiple times, the probability that any particular person polled disapproves of the president is p, and $d \sim \text{Binomial}(p, n)$, since people's opinions are independent and each person polled has chance p of being among those who disapprove of the president. Then, $E[d] = pn$, so that, by Hoeffding's Inequality, for $\epsilon = 2\% = 0.02$, $19/20 = P((p-\epsilon)n \leq d \leq (p+\epsilon)n) \geq 2e^{-2\epsilon^2 n}$, and thus. Solving for n gives

$$n \geq -\frac{\ln(19/40)}{2\epsilon^2} \approx \boxed{931.}$$