MATH 5763 Homework 7 Due Fri, 4/1/11

Problem 1. [**Do NOT turn it in, just read it and think about it!**]
A sequence $\mathbf{a} = \{a_n\}_{n=0}^{\infty}$ of real numbers can contain a lot of information. One concise way of storing this information is to wrap up the numbers a_n together in a "generating function". For example, let us define the (ordinary) *generating function* of the sequence $\mathbf{a}$ in the function $G_{\mathbf{a}}$ defined by

$$G_{\mathbf{a}}(s) = \sum_{n=0}^{\infty} a_n s^n \quad \text{for those } s \in \mathbb{R} \text{ for which the sum converges .}$$

The sequence $\mathbf{a}$ may be reconstructed from the function $G_{\mathbf{a}}$ by setting $a_n = \frac{1}{n!} G_{\mathbf{a}}^{(n)}(0)$, where $f^{(n)}$ denotes the nth derivative of the function f. Generating functions are considered in many books on combinatorics, discrete mathematics, and probability (among others); a readable and freely available book is Herbert Wilf's *generatingfunctionology* (its second edition is freely available at `www.math.upenn.edu/~wilf/DownldGF.html`).
The *convolution* of the sequences $\mathbf{a} = \{a_n\}_{n=0}^{\infty}$ and $\mathbf{b} = \{b_n\}_{n=0}^{\infty}$ is the sequence $\mathbf{c} = \{c_n\}_{n=0}^{\infty}$ defined by

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \quad \left(= \sum_{k=0}^{n} a_{n-k} b_n \right) .$$

Sometimes the convolution of $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathbf{a} * \mathbf{b}$.

(a) Let $z_n = (\cos\theta + \mathrm{i}\sin\theta)^n$, where $\mathrm{i} = \sqrt{-1}$, and θ is a fixed real number. Show that the generating function of the sequence $\mathbf{z} = \{z_n\}_{n=0}^{\infty}$ is

$$G_{\mathbf{z}}(s) = \frac{1}{1 - s(\cos\theta + \mathrm{i}\sin\theta)} \quad \text{for } \; |s| < 1 .$$

(b) Prove that, if $\mathbf{a}$ and $\mathbf{b}$ have generating functions $G_{\mathbf{a}}$ and $G_{\mathbf{b}}$, then the generating function of $\mathbf{c} = \mathbf{a} * \mathbf{b}$ is $G_{\mathbf{c}}(s) = G_{\mathbf{a}}(s)\, G_{\mathbf{b}}(s)$.

(c) Obtain the combinatorial identity $\displaystyle\sum_{k=0}^{N} \binom{N}{n}^2 = \binom{2N}{N}$, where $N \in \mathbb{N}$, by noticing that its left-hand side can be thought of as the convolution of the sequence

$$a_n = \begin{cases} \binom{N}{n} & \text{for } 0 \le n \le N \\ 0 & \text{for } n \ge N+1 \end{cases}$$

with itself, and using the fact proved in part (b) about $G_{\mathbf{a}*\mathbf{a}}$.

(d) It is a well-known fact from elementary probability (and an easy exercise for you) that if X and Y are independent random variables taking values in $\mathbb{Z}_+ = \{0, 1, 2, 3, \ldots\}$, then the probability mass function of their sum, $Z = X + Y$, is the convolution of the probability mass functions of X and Y: $p_Z = p_X * p_Y$. Here we think of a p.m.f. p_X as a sequence $\mathbf{a}$ with $a_n = p_X(n) = \mathbb{P}(X = n)$.

Let X and Y be independent Poisson random variables with parameters λ and μ, respectively. Compute explicitly the probability generating functions

$$G_X(s) = \sum_{n=0}^{\infty} p_X(n)\, s^n \quad \text{and} \quad G_Y(s) = \sum_{n=0}^{\infty} p_Y(n)\, s^n \ ,$$

and use them to show that $X+Y$ is a Poisson random variable and to find the parameter of the distribution of $X + Y$.

Problem 2. The *(probability) generating function* of a random variable X taking values in $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ is defined to be the generating function of its probability mass function p_X:

$$G_X(s) = \mathbb{E}[s^X] = \sum_{n=0}^{\infty} p_X(n)\, s^n \quad \left(= \sum_{n=0}^{\infty} \mathbb{P}(X = n)\, s^n \right) .$$

(a) Let X_1, X_2, ... be a sequence of independent identically distributed random variables (with values in $\mathbb{Z}_+$) with common generating function G_X. Let N be a random variable taking values in $\mathbb{Z}_+$ which is independent of the X_j's, and let G_N be the generating function of N. Define the random variable Y_N by $Y_N = X_1 + X_2 + \cdots + X_N$ if $N \geq 1$ and $Y_N = 0$ if $N = 0$. Show that Y_N has generating function given by $G_{Y_N}(s) = G_N(G_X(s))$.

(b) Let Z be a Poisson random variable with parameter Λ, where Λ is a Poisson random variable with parameter μ. Compute G_Z and $\mathbb{E}[Z]$.
Hint: The generating function of $\Lambda \sim \text{Poisson}(\mu)$ is $G_\Lambda(s) = \mathrm{e}^{\mu(s-1)}$.

(c) Let V be a Poisson random variable with parameter Θ, where Θ is an exponential random variable with parameter ν. Show that $V + 1$ has a geometric distribution, and find the parameter of this distribution by using two methods:

- direct computation of the p.m.f. of $V + 1$;
- computing the generating function G_V or G_{V+1} (and then convincing me that your result indeed implies the desired conclusion).

Hint: If $\Theta \sim \text{Exponential}(\nu)$, then $f_\Theta(x) = \nu \mathrm{e}^{-\nu x} \chi_{[0,\infty)}(x)$; if $W \sim \text{Geometric}(p)$, then $p_W(n) = (1-p)^{n-1} p$ for $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$, and $G_W(s) = \dfrac{ps}{1 - (1-p)s}$.

Problem 3. A coin is tossed repeatedly, and on each individual toss heads turns up with probability $p \in (0,1)$ (and tails with probability $1-p$). Let h_n be the probability of an even number of heads in the first n tosses, with the convention that 0 is an even number:

$$h_n = \mathbb{P}(0 \text{ or } 2 \text{ or } 4 \text{ or } 6 \text{ or } \ldots \text{ heads in the first } n \text{ tosses}) .$$

(a) What is h_0 and why?

(b) Show that for $n \in \mathbb{N}$, the numbers h_n satisfy the following difference equation:

$$h_n = (1-p)h_{n-1} + p(1-h_{n-1}) .$$

Hint: Condition on the outcome of the first toss. In other words, use that the events {heads on the first toss} and {tails on the first toss} form a partition of the sample space, and use the law of total probability.

(c) Deduce that the generating function of the sequence $\mathbf{h} = \{h_n\}_{n=0}^{\infty}$ is

$$G_{\mathbf{h}}(s) = \frac{1}{2}\left(\frac{1}{1+2ps-s} + \frac{1}{1-s}\right) .$$

Hint: Multiply the identity derived in (b) by s^n and sum over $n = 1, 2, 3, \ldots$ to show that $G_{\mathbf{h}}(s) - 1 = (1-2p)sG_{\mathbf{h}}(s) + \frac{ps}{1-s}$.

(d) Use the formula for the sum of a geometric series to compute h_n from $G_{\mathbf{h}}$.

(e) Obtain h_n by solving the difference equation from (b) directly.

Hint: The equation in (b) can be written in the form of a linear nonhomogeneous recurrence relation: $h_n = (1-2p)h_{n-1} + p$. One way to solve it is to write this relation and the relation following from it by replacing n by $n+1$, then subtract the two equations to derive a homogeneous recurrence relation involving h_{n+1}, h_n, and h_{n-1}, and use the methods described in Problem 5 of Homework 3.

Problem 4. Let $X_1, X_2, \ldots$ be a sequence of independent Uniform$(0,1)$ random variables. Let $t \in (0,1]$, and

$$N(t) := \min\left\{ n \in \mathbb{N} \;:\; \sum_{k=1}^{n} X_k > t \right\}$$

be the smallest number of X's that need to be added so that their sum exceeds t, as shown in Figure 1 (where $N(t) = 4$). Define $M(t) := \mathbb{E}[N(t)]$. From the definition, it is clear that $N(t)$ is a random variable taking values in $\mathbb{N}$.

(a) Explain why, for any $x_1 \in (0,1]$,

$$E\big[N(t)\big|X_1 = x_1\big] = \begin{cases} 1 & \text{for } t < x_1 , \\ 1 + M(t-x_1) & \text{for } x_1 \le t . \end{cases}$$

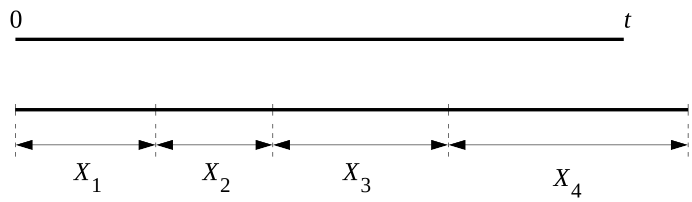

Figure 1: Covering the interval $[0, t]$ with lines with random lengths X_m.

Hint: This question does not require any calculation! Draw a picture and think about the following questions. If $t < x_1 = X_1$, how many X's do you need in order to cover the interval $[0, t]$? If $X_1 = x_1 \leq t$, then how many X's do you need in addition to X_1 in order to cover $[0, t]$?

(b) Show that $M(t)$ satisfies the integral equation

$$M(t) = 1 + \int_0^t M(t - x_1)\, \mathrm{d}x_1 \ . \tag{1}$$

Hint: Use the Tower Rule by conditioning on X_1. Don't forget that $t \in (0, 1]$.

(c) Solve the integral equation for $M(t)$ by first converting it to a differential equation. First change variables in the integral to rewrite (1) in the form $M(t) = 1 + \displaystyle\int_0^t M(y)\, \mathrm{d}y$ (write explicitly the change of variables). Differentiate both sides of this equation to derive a *differential* equation for $M(t)$. What is the initial condition for $M(0)$ that the function $M(t)$ should satisfy (justify your answer with a couple of sentences). Solve the initial value problem for $M(t)$ you have obtained.

(d) Now solve the integral equation for $M(t)$ by using Laplace transform. You can use the following facts about Laplace transform, without proving them: if $h(x) = x^n$, for $n \in \mathbb{N}$, then $\mathcal{L}[h](\xi) = \frac{n!}{\xi^{n+1}}$, $\xi > 0$; if $h(x) = \mathrm{e}^{ax}$ (for any $a \in \mathbb{R}$), then $\mathcal{L}[h](\xi) = \frac{1}{\xi - a}$, $\xi > a$; if $A(t) = \int_0^t M(t - x_1)\, f(x_1)\, \mathrm{d}x_1$, then $\mathcal{L}[A](\xi) = \mathcal{L}[m](\xi)\, \mathcal{L}[f](\xi)$.

Problem 5. The toll collected from the traffic passing through the toll booth on Highway 44 between Oklahoma City and Tulsa can be modeled for the hours between 9 a.m. and 5 p.m. by a compound Poisson process. Assume that the toll booth serves the arriving vehicles instantaneously, so that there are no waiting lines.

The vehicles can be divided into two big categories – personal vehicles and commercial vehicles. The personal vehicles arrive at the toll booth with average frequency 7 personal vehicles per minute, while the commercial vehicles arrive with average frequency 3 commercial vehicles per minute.

There are three types of personal vehicles – 80 % of the personal vehicles are cars, 15 % are SUVs and 5 % are RVs. There are four types of commercial vehicles – pick-up trucks, normal-size trucks, 18-wheelers, and busses; the probability with which a commercial vehicle belongs to each of these four types is 40 %, 30 %, 20 %, and 10 %, respectively.

The toll rates are the following: car \$1, SUV \$3, RV \$5; pick-up truck \$3, normal-size truck \$5, 18-wheeler \$8, bus \$10.

Please answer the questions below. Define clearly your notations, and use the concrete numbers given in this problem. You are allowed to use the theoretical results derived in class, but please write explicitly what results you use.

(a) Think of the toll collected from the personal vehicles as a compound Poisson process $Y_1(t)$, where at each arrival of a personal vehicle the collected toll is random. What is the p.m.f. of the random variable describing the collected toll from a personal vehicle? What is the rate of the Poisson process describing the moments of arrival of personal vehicles?

(b) Find the moment generating function of the process $Y_1(t)$, and use your result to find the average value of the toll collected in a period of 1 hour, and the variance of this toll.

(c) Answer the same questions as in part (a) about the Poisson process $Y_2(t)$ describing the toll collected from the commercial vehicles.

(d) Answer the same questions as in (b), but for the process $Y_2(t)$.

(e) Define the random process $Y(t) = Y_1(t) + Y_2(t)$ of the toll collected from all vehicles passing through the toll booth. We can think of this random process as a compound Poisson process. What is the frequency of the Poisson process with which the events of this random process occur? What is the p.m.f. of the toll collected from each vehicle passing through the toll booth (without making a distinction between personal and commercial vehicles)?

(f) Write explicitly the moment generating function of the random process $Y(t)$, as well as $\mathbb{E}[Y(t)]$ and the variance of $Y(t)$. On average, how much toll will be collected from 10 a.m. to 11 a.m.?

Problem 6. The arrival of customers through the Walmart in Norman near I-35 can be thought as a (simple, not compound) Poisson process with a time-dependent rate $\lambda(t)$. If we measure the time of the day in hours, starting at midnight, then the rate of the process can be modeled by the function

$$\lambda(t) = 500 \left(1 - \sin\frac{2\pi t}{24}\right) \quad \text{persons per hour} \ .$$

Let Y be the number of customers arriving at the store between 2 p.m. and 5 p.m. Find the average value and the variance of this random variable.