Homework #6, due in class on Thursday, March 17th

Suggested reading:

1. For this assignment: Poisson approximation to binomial (end of 2.2 in BT, see also 2.3 in Durrett), independent random variables (2.7 in BT, 3.1 in Pitman and very briefly 3.4 in Durrett) and conditional PMF and conditional expectations (2.6 in BT, 6.1 in Pitman and very briefly 3.4 in Durrett).

2. For next week's classes: BT, §3.1 and 3.2; Pitman, §4.1; Durrett §5.1 and 5.2

General convention for this assignment: All random variables considered in this assignments are assumed to have range inside $\mathbb{Z}_{\geq 0}$ (non-negative integers); in particular, they are all discrete.

Problems:

The first four problems in this assignment deal with the notion of a *probability generating function of a random variable* (called simply a generating function below). So let X be a random variable with $R(X) \subseteq \mathbb{Z}_{\geq 0}$, and define its generating function G_X by

$$G_X(t) = \sum_{k \in R(X)} P(X = k) \cdot t^k.$$

Here t is a formal variable, so $G_X(t)$ is defined as the sum of a power series in t. By basic properties of series, this power series converges at least when $-1 < t \leq 1$, so G_X is a function defined at least on the interval $(-1, 1]$. The generating function G_X becomes a useful tool when one can compute an explicit (closed) formula for it, which is the case for many important distributions.

For instance, suppose that X is a geometric random variable with parameter p, and let $q = 1 - p$. Then

$$G_X(t) \;=\; \sum_{k=1}^{\infty} pq^{k-1}t^k \;=\; \frac{p}{q}\sum_{k=1}^{\infty}(qt)^k \;=\; \frac{p}{q} \cdot \frac{qt}{1-qt} \;=\; \frac{pt}{1-(1-p)t}.$$

Problem 1:

(a) Let X be a Poisson random variable with parameter λ. Prove that $G_X(t) = e^{\lambda(t-1)}$.

(b) Let X be a binomial (n, p) random variable (n is the number of trials and p is the probability of success). Prove that

$$G_X(t) = (1-p)^n \left(1 + \frac{pt}{1-p}\right)^n.$$

Problem 2: Let X and Y be independent random variables.

(a) Prove that for every $n \in \mathbb{Z}_{\geq 0}$ we have

$$P(X+Y=n) = \sum_{k=0}^{n} P(X=k)P(Y=n-k).$$

(b) Use (a) to prove that $G_{X+Y}(t) = G_X(t)G_Y(t)$. **Hint:** When you multiply two power series in t, how is coefficient of t^n obtained?

Problem 3: Let X and Y be independent random variables.

(a) Assume that X is Poisson with parameter λ and Y is Poisson with parameter μ. Use Problems 1 and 2 to find PMF of $X + Y$.

(b) Now assume that X is binomial with parameters (n, p) and Y is binomial with parameters (m, p) (the value of p is the same for both X and Y). Find the PMF of $X + Y$ in two different ways – using problem 1 and 2 and then by using interpretation of binomial random variables as the number of successes (naturally you should get the same answer).

Problem 4: This problem shows how generating functions can be used to compute mean and variance. Let X be a random variable. Starting with equality $G_X(t) = \sum\limits_{k \in R(X)} P(X=k)t^k$, differentiating both sides and then plugging in $t = 1$, we first get $G'_X(t) = \sum\limits_{k \in R(X)} k \cdot P(X=k)t^{k-1}$ and then $G'_X(1) = \sum\limits_{k \in R(X)} k \cdot P(X=k)$, so $EX = G'_X(1)$.

(a) Use a similar trick to show that $Var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$

(b) Use the above formulas and the result of problem 1 to compute the mean and the variance for the binomial and Poisson distributions.

Problem 5: Suppose we roll a fair (6-sided) die twice, let X_1 be the number on the first roll and X_2 the number on the second roll, $X = \max\{X_1, X_2\}$ and $Y = \min\{X_1, X_2\}$.

(a) For every pair of integers a and b between 1 and 6 compute $P(X = a|Y = b)$ and $P(X = b|Y = a)$.

(b) Now for each integer a between 1 and 6 compute $E[X|Y=a]$ (the answer should be a function of a; you probably have to do it separately for each a first, but in the end the pattern should be pretty clear) and $E[Y|X=a]$

(c) Now repeat (b) for an n-sided die (with n arbitrary).

Problem 6: A manufacturer sells certain items in boxes, with 100 items per box. It is known that $k\%$ of all items are defective, where k is an integer, and that approximately 95% of all boxes contain at least one defective item. Based on Poisson approximation, what is the most likely value of k? You do not need a calculator, but you have to use the fact that e is approximately 2.7.