281[1] *Let* $G(s) = \sum_{k=0}^{\infty} f_X(k)s^k$ *where* $f_X(k) = \mathbf{P}(X = k)$ *for* $k \geq 0$. *Show that:*

(a) $\displaystyle\sum_{k=0}^{\infty} \mathbf{P}(X < k)s^k = \frac{sG(s)}{1-s}$.

(b) $\displaystyle\sum_{k=0}^{\infty} \mathbf{P}(X \geq k)s^k = \frac{1-sG(s)}{1-s}$.

(a.) Here are three ways to see this formula. The first one uses the Cauchy Product Formula from Calculuc II:

$$\left(\sum_{i=0}^{\infty} a_i s^i\right)\left(\sum_{j=0}^{\infty} b_j s^j\right) = \sum_{k=0}^{\infty} c_k s^k$$

where

$$c_k = \sum_{i=0}^{k} a_i\, b_{k-i}$$

From the hint, because X is nonnegative integer valued, since $\mathbf{P}(X < 0) = 0$ and by switching indices to $j = k - 1$,

$$H(s) = \sum_{k=0}^{\infty} \mathbf{P}(X < k)s^k = s\sum_{k=1}^{\infty} \mathbf{P}(X \leq k-1)s^{k-1} = s\sum_{j=0}^{\infty} \mathbf{P}(X \leq j)s^j.$$

Theses terms can be recognized as the coefficients of the product

$$c_k = \mathbf{P}(X \leq k) = \sum_{i=0}^{k} \mathbf{P}(X = i) = \sum_{i=0}^{k} a_i b_{k-i}$$

where $a_i = \mathbf{P}(X = i)$ and $b_i = 1$ for all $i \geq 0$. Hence, by the Cauchy Product formula,

$$s \cdot G(s) \cdot \frac{1}{1-s} = s\left(\sum_{i=0}^{\infty} \mathbf{P}(X = i)s^i\right)\left(\sum_{j=0}^{\infty} s^j\right) = s\left(\sum_{k=0}^{\infty} c_k s^k\right) = H(s).$$

The second way is to notice that $c_k - c_{k-1} = \mathbf{P}(X \leq k) - \mathbf{P}(X \leq k-1) = \mathbf{P}(X = k) = f_X(k) = a_k$ for all $k \geq 0$ and $c_0 = a_0$. Put $c_{-1} = 0$. Multiplying by s^{k+1} and summing,

$$(1-s)H(s) = H(s) - sH(s) = s\sum_{k=0}^{\infty} c_k s^k - s^2\sum_{k=0}^{\infty} c_{k-1}s^{k-1} = s\sum_{k=0}^{\infty} a_k s^k = sG(s).$$

A third way was discovered by some clever students in our class. The idea is to switch the order of summation, thus

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbf{P}(X < k)s^k &= \sum_{k=1}^{\infty} \mathbf{P}(X \le k-1)s^k \\
&= \sum_{k=1}^{\infty}\sum_{j=0}^{k-1} \mathbf{P}(X = j)s^k \\
&= \sum_{j=0}^{\infty} \mathbf{P}(X = j) \sum_{k=j+1}^{\infty} s^k \\
&= \sum_{j=0}^{\infty} \mathbf{P}(X = j) \left(\frac{s^{j+1}}{1-s}\right) \\
&= \frac{s}{1-s} \sum_{j=0}^{\infty} \mathbf{P}(X = j)s^j \\
&= \frac{sG(s)}{1-s}.
\end{aligned}$$

(b.) The second formula follows from $\mathbf{P}(X \ge k) = 1 - \mathbf{P}(X < k)$. Multiplying by s^k, summing and using part (a.),

$$\sum_{k=0}^{\infty} \mathbf{P}(X \ge k)s^k = \sum_{k=0}^{\infty} s^k - \sum_{k=0}^{\infty} \mathbf{P}(X < k)s^k = \frac{1}{1-s} - H(s) = \frac{1-sG(s)}{1-s}.$$

A*. *Consider the coupon collecting problem with three types of coupons. Let T be the number of boxes needed until you first possess all three types. Find $\mathbf{P}(T = k)$, $\mathbf{E}(T)$ and $\mathbf{Var}(T)$ using the probability generating function.*

The first coupon comes in the first box. Let X_1 be the further number of boxes until a different coupon appears. Thus X_1 is a geometric variable with probability of success $p = \frac{2}{3}$. The generating function for X_1 is

$$G_1(s) = \frac{ps}{1-qs} = \frac{2s}{3-s}.$$

Let X_2 be the further number of boxes until the last type of coupon appears. X_2 is a geometric variable with probability if success $p = \frac{1}{3}$. Thus

$$G_2(s) = \frac{ps}{1-qs} = \frac{s}{3-2s}.$$

The number of boxes to get all three types is $T = 1 + X_1 + X_2$. Since X_1 and X_2 are independent, the generating function of the sum is the product of pgf's. Adding one multiplies the pgf by s. Hence

$$G(s) = sG_1(s)G_2(s) = \frac{2s^3}{(3-2s)(3-s)}.$$

To get the power series for $G(s)$, it helps to cleave it into easier pieces using partial fractions. Since the denominator is factored into distinct linear factors, these occur as denominators of simpler terms with unknown constants A and B

$$\frac{2}{(3-2s)(3-s)} = \frac{A}{3-2s} + \frac{B}{3-s}.$$

To solve for the constants, multiply the denominators

$$2 = A(3-s) + B(3-2s).$$

Taking $s = 3$ we find $B = -\frac{2}{3}$. Taking $s = 0$ we find $A = \frac{4}{3}$. Adding, and simplifying denominators

$$G(s) = \frac{\frac{4}{3}s^3}{3-2s} - \frac{\frac{2}{3}s^3}{3-s} = \frac{\frac{4}{9}s^3}{1-\frac{2}{3}s} - \frac{\frac{2}{9}s^3}{1-\frac{1}{3}s}.$$

Finally, we use the geometric series, and substituting $k = j+3$,

$$G(s) = \frac{4}{9}\sum_{j=0}^{\infty}\left(\frac{2}{3}\right)^j s^{j+3} - \frac{2}{9}\sum_{j=0}^{\infty}\left(\frac{1}{3}\right)^j s^{j+3} = \sum_{k=3}^{\infty}\left(\frac{2^{k-1}-2}{3^{k-1}}\right)s^k.$$

It follows that, for $k \ge 3$,

$$\mathbf{P}(T=k) = f_T(k) = \frac{2^{k-1}-2}{3^{k-1}}.$$

Assuming the derivatives exist, they give the first and second moments. Taking logarithmic derivatives

$$\frac{G'(s)}{G(s)} = \frac{3}{s} + \frac{2}{3-2s} + \frac{1}{3-s},$$

$$\frac{G''(s)}{G(s)} - \frac{\left(G'(s)\right)^2}{G(s)^2} = -\frac{3}{s^2} + \frac{4}{(3-2s)^2} + \frac{1}{(3-s)^2}.$$

So, using the property of pgf's that $G(1) = 1$,

$$G'(1) = \frac{G'(1)}{G(1)} = 3 + 2 + \frac{1}{2} = \frac{11}{2},$$

$$G''(1) - \left(G'(1)\right)^2 = \frac{G''(1)}{G(1)} - \frac{\left(G'(1)\right)^2}{G(1)^2} = -3 + 4 + \frac{1}{4} = \frac{5}{4}.$$

The expectation and variance are

$$\mathbf{E}(T) = G'(1) = \frac{11}{2} = 5\tfrac{1}{2},$$

$$\mathbf{Var}(T) = G''(1) - \left(G'(1)\right)^2 + G'(1) = \frac{5}{4} + \frac{11}{2} = 6\tfrac{3}{4}.$$

B*. *Flip a fair coin repeatedly until you get two consecutive heads. This takes X flips. Derive the pgf. Use it to find the expectation and the variance of X.*

Suppose $\mathbf{P}(\text{head on } n\text{th spin}) = p$ (for us $p = \frac{1}{2}$.) Following the argument in the text, let's condition on the first couple of tosses. If the first toss is tail, then the conditional expectation of the number of tosses is the same as at the outset, except that one more toss (the first) was used. Thus

$$\mathbf{E}(s^X \mid T_1) = \mathbf{E}(s^{X+1}) = s\,\mathbf{E}(s^X).$$

If the first two tosses are a head then tail, then the conditional expectation of the number of tosses is the same as at the outset, except that two more tosses were used. Thus

$$\mathbf{E}(s^X \mid H_1T_2) = \mathbf{E}(s^{X+2}) = s^2\,\mathbf{E}(s^X).$$

If the first two tosses are a heads then two consecutive heads have occurred and $X = 2$. Thus

$$\mathbf{E}(s^X \mid H_1H_2) = \mathbf{E}(s^2) = s^2.$$

Using the partitioning formula for expectation,

$$\begin{aligned}\mathbf{E}(s^X) &= \mathbf{E}(s^X \mid T_1)\,\mathbf{P}(T_1) + \mathbf{E}(s^X \mid H_1T_2)\,\mathbf{P}(H_1T_2) + \mathbf{E}(s^X \mid H_1H_2)\,\mathbf{P}(H_1H_2)\\ &= s\,\mathbf{E}(s^X)q + s^2\,\mathbf{E}(s^X)pq + s^2p^2.\end{aligned}$$

Solving for $\mathbf{E}(s^X)$ we get

$$G_X(s) = \mathbf{E}(s^X) = \frac{p^2s^2}{1 - qs - pqs^2}.$$

Taking derivatives of the logarithm we see that

$$\begin{aligned}\frac{G'_X(s)}{G_X(s)} &= \frac{2}{s} + \frac{q + 2pqs}{1 - qs - pqs^2},\\ \frac{G''_X(s)}{G_X(s)} - \frac{\big(G'_X(s)\big)^2}{G_X(s)^2} &= -\frac{2}{s^2} + \frac{2pq}{1 - qs - pqs^2} + \frac{(q + 2pqs)^2}{\big(1 - qs - pqs^2\big)^2}.\end{aligned}$$

So, using the property of pgf's that $G(1) = 1$,

$$\begin{aligned}G'(1) &= \frac{G'(1)}{G(1)} = 2 + \frac{q + 2pq}{1 - q - pq} = \frac{1 + p}{p^2},\\ G''(1) - \big(G'(1)\big)^2 &= \frac{G''(1)}{G(1)} - \frac{\big(G'(1)\big)^2}{G(1)^2} = -2 + \frac{2pq}{1 - q - pq} + \frac{(q + 2pq)^2}{\big(1 - q - pq\big)^2}\\ &= -2 + \frac{2q}{p} + \frac{q^2(1 + 2p)^2}{p^4}.\end{aligned}$$

Thus the expectation and variance for $p = \frac{1}{2}$ are

$$\begin{aligned}\mathbf{E}(T) &= G'(1) = 6,\\ \mathbf{Var}(T) &= G''(1) - \big(G'(1)\big)^2 + G'(1) = 16 + 6 = 22.\end{aligned}$$