ST202/ST206 – Michaelmas Term

Solutions to problem set 6

1. (a) The transformation $g(x) = 1/x$ is strictly monotonic for $x > 0$, so we can apply the change-of-variable formula. Using the fact that $g^{-1}(y) = 1/y$, we have

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = f_X(1/y) \left| \frac{d}{dy} \frac{1}{y} \right| = f_X(1/y) \frac{1}{y^2} \,,$$

for $y > 0$, and $f_Y(y) = 0$ otherwise.

(b) If $X \sim \text{Exp}(\lambda)$, applying the formula from the first part gives

$$f_Y(y) = f_X(1/y) \frac{1}{y^2} = \lambda y^{-2} e^{-\lambda/y} \,, \quad \text{for } y > 0.$$

2. First, notice that $S_X(x-) = 1 - F_X(x-)$, which is equal to $1 - F_X(x)$ if X is continuous. In the discrete case, we need to be a bit more careful.

(a) We know that $F_X(x) = 1 - e^{-\lambda x}$, so $S_X(x) = e^{-\lambda x}$. The hazard function is, thus,

$$\lambda_X(x) = \frac{f_X(x)}{S_X(x)} = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda \,.$$

(b) Here X is discrete and integer-valued, so if x is a positive integer we have $F_X(x-) = P(X < x) = F_X(x-1)$. We worked out in Problem Set 3 that $F_X(x) = 1 - (1-p)^x$, thus,

$$\lambda_X(x) = \frac{f_X(x)}{1 - F_X(x-1)} = \frac{p(1-p)^{x-1}}{(1-p)^{x-1}} = p \,.$$

(c) In Problem Set 4 we showed that $F_X(x) = 1 - (1+x)^{-a+1}$, so

$$\lambda_X(x) = \frac{f_X(x)}{S_X(x)} = \frac{(a-1)(1+x)^{-a}}{(1+x)^{-a+1}} = \frac{a-1}{x+1} \,.$$

(d) Also from Problem Set 4, $F_X(x) = 1 - e^{-cx^\tau}$, which gives

$$\lambda_X(x) = \frac{f_X(x)}{S_X(x)} = \frac{c\tau x^{\tau-1} e^{-cx^\tau}}{e^{-cx^\tau}} = c\tau x^{\tau-1} \,.$$

The Exponential and Geometric distributions have constant hazard, which is a consequence of the memoryless property (more on this later). The Pareto has decreasing hazard. The Weibull can have decreasing, constant, or increasing hazard, depending on the value of τ.

3. (a) The monkey will get each letter (the first, the second, ..., the k^{th}) correct with probability 1/26. As the keystrokes are independent, the probability all k letters are correct is the product of these probabilities, $(1/26)^k$.

   (b) Suppose that we check the output every k keystrokes. Let E_n, for $n = 1, 2, \ldots$, be the event that keystrokes $(n-1)k+1, \ldots, nk$ produce the full works of Shakespeare. In this notation, the event we considered in part (a) is E_1.

   Now consider the sequence of events $E_1, E_2, \ldots$, which all have the same probability. Since

   $$\sum_{n=1}^{\infty} P(E_n) = \sum_{n=1}^{\infty} 26^{-k} = \infty\,,$$

   and the $\{E_n\}$ are independent, the second Borel Cantelli lemma tells us that infinitely many of them occur with probability 1. In other words, the monkey will produce infinitely many copies of the works of Shakespeare with probability 1!

   Notice that this is a conservative scheme; if the monkey produces the desired output starting from, say, the second keystroke, we will not count it as correct.

   (c) No. In part (a), if the probability that the j^{th} letter is correct is p_j, we have $P(E_1) = \prod_{j=1}^{k} p_j$. As long as this is positive, i.e. each keystroke is correct with some positive probability (however small), the sum in part (b) will diverge.

   (d) Yes; we need independence or the conditions of the lemma are not met. For example, consider the case where the monkey chooses the first key at random, then continues to hit the same key with every subsequent stroke.

4. (a) Random variables with the same cumulant- (or moment-) generating function have the same distribution, so we would conclude that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$, i.e. X_n converges to X in distribution.

   (b) Recall that, if $U \sim \text{Pois}(\lambda)$, its MGF is $M_U(t) = \exp\left\{\lambda(e^t - 1)\right\}$.

We thus have

$$
\begin{aligned}
K_{X_n}(t) &= \log M_{X_n}(t) = \log \mathbb{E}(e^{tX_n}) = \log \mathbb{E}(e^{t(Y_n - n)/\sqrt{n}}) \\
&= \log \left\{ \mathbb{E}(e^{\frac{t}{\sqrt{n}} Y_n}) e^{-t\sqrt{n}} \right\} = \log M_{Y_n}(t/\sqrt{n}) - t\sqrt{n} \\
&= n(e^{t/\sqrt{n}} - 1) - t\sqrt{n} \\
&= n \left[\frac{t}{\sqrt{n}} + \frac{1}{2!} \left(\frac{t}{\sqrt{n}} \right)^2 + \frac{1}{3!} \left(\frac{t}{\sqrt{n}} \right)^3 + \dots \right] - t\sqrt{n} \\
&= \left[t\sqrt{n} + \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}} + \dots \right] - t\sqrt{n} \\
&= \frac{t^2}{2} + \frac{t^3}{3!\sqrt{n}} + \dots \\
&\to \frac{t^2}{2} \text{ as } n \to \infty
\end{aligned}
$$

because all the other terms have a positive power of n in the denominator and go to 0. We identify this as the cumulant-generating function of a standard normal, i.e. $X_n \xrightarrow{d} Z$, where $Z \sim \mathrm{N}(0, 1)$.

For large n, if the distribution of X_n is approximately standard normal, we conclude that the distribution of $Y_n = \sqrt{n} Y_n + n$ must be approximately $\mathrm{N}(n, n)$. This is the normal approximation to the Poisson distribution.