Lecture 16, Feb 15, 2023

Memoryless Property of the Exponential Distribution

- Suppose we have a random variable X with exponential distribution $f(x) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$
- Consider $P(X \geq s+t | X \geq s)$, i.e. if there has been no event for time s, what is the probability that there are no more events for another time t?
  - $$\begin{aligned} P(X \geq s+t | X \geq s) &= \frac{P(X \geq s+t \cap X \geq s)}{P(X \geq s)} \\ &= \frac{P(X \geq s+t)}{P(X \geq s)} \\ &= \frac{\int_{s+t}^{\infty} \frac{1}{\beta}e^{-\frac{x}{\beta}}\,\mathrm{d}x}{\int_{s}^{\infty} \frac{1}{\beta}e^{-\frac{x}{\beta}}\,\mathrm{d}x} \\ &= \frac{e^{-\frac{(s+t)}{\beta}}}{e^{-\frac{s}{\beta}}} \\ &= e^{-\frac{t}{\beta}} \\ &= P(X \geq t) \end{aligned}$$
- This means it doesn't matter how long we've waited – the past has no impact on the probability distribution in the future
  - This is known as the *memoryless* property
- Note that this is a modelling assumption that we have to be aware of; it is not a statement about reality

Functions of Random Variables

- Suppose X is a discrete random variable with PMF $f(x)$; suppose $Y = u(x)$ where u is one-to-one (aka bijective, invertible); what is the PMF of Y?
  - $X = u^{-1}(Y)$
  - $g(y) = P(Y = y)$ since the RVs are discrete
  - $g(y) = P(u^{-1}(Y) = u^{-1}(y)) = P(X = u^{-1}(y)) = f(u^{-1}(y))$
- If X is a discrete random variable with PDF $f(x)$, let $g(y)$ and $G(y)$ be the PDF and CDF of Y
  - We can no longer do the same thing as in the discrete case because $P(Y = y) = 0$, so we must consider the CDF
  - $G(y) = P(Y \leq y) = P(u^{-1}(Y) \leq u^{-1}(y)) = P(X \leq u^{-1}(y)) = \displaystyle\int_{-\infty}^{u^{-1}(y)} f(t)\,\mathrm{d}t$
  - To get back the PDF we need to differentiate
  - $g(y) = \dfrac{\mathrm{d}G}{\mathrm{d}y} = \dfrac{\mathrm{d}}{\mathrm{d}y}\displaystyle\int_{-\infty}^{u^{-1}(y)} f(t)\,\mathrm{d}t = f(u^{-1}(y))\dfrac{\mathrm{d}u^{-1}(y)}{\mathrm{d}y}$ using Leibniz's integral rule
  - Note here we made the assumption that $u(y)$ is strictly increasing; if it's strictly decreasing, then we need to flip the inequality in $G(y)$
    - This means we need to add an absolute value around $\dfrac{\mathrm{d}u^{-1}}{\mathrm{d}y}$

Summary

Given a discrete random variable X with PMF $f(x)$, if $Y = u(x)$ where u is invertible, then the PMF of Y is given by

$$g(y) = f(u^{-1}(y))$$

If X is continuous, then we have

$$g(y) = f(u^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}u^{-1}(y)\right|$$

Functions of Multiple Random Variables and Non-Invertible Functions (Textbook Ch 7)

Theorem

Let X_1, X_2 be two discrete random variables with joint distribution $f(x_1, x_2)$; let

$$Y_1 = u_1(X_1, X_2), Y_2 = u_2(X_1, X_2)$$

define a one-to-one transformation, such that we may find unique inverse functions

$$x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$$

; then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f(w_1(y_1, y_2), w_2(y_1, y_2))$$

- If we just want the distribution of some $Y_1 = u_1(X_1, X_2)$, then we can make up a Y_2 such that we have a one-to-one transformation; this allows us to get $g(y_1, y_2)$, from which we can find $g(y_2) = \sum_{y_2} g(y_1, y_2)$
  - e.g. if $Y_1 = X_1 + X_2$ then we can let $Y_2 = X_2$, then our inverse functions are given by $x_1 = y_1 - y_2, x_2 = y_2$

Theorem

Let X_1, X_2 be two continuous random variables with joint distribution $f(x_1, x_2)$; let $Y_1 = u_1(X_1, X_2), Y_2 = u_2(X_1, X_2)$ define a one-to-one transformation (like in the discrete case) with $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$, then the joint probability distribution of Y_1 and Y_2 is given by

$$g(y_1, y_2) = f(w_1(y_1, y_2), w_2(y_1, y_2))|J|$$

where J is the Jacobian given by

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

- This theorem is a generalization of the one variable function of a continuous RV; in the case of a single RV, the Jacobian becomes simply the derivative of the inverse function u^{-1}

Theorem

Let X be a continuous random variable with distribution $f(x)$; let $Y = u(X)$ define a transformation such that u is not one-to-one, but the interval over which X is defined can be partitioned into k mutually disjoint sets such that each of the inverse functions

$$x_1 = w_1(y), x_2 = w_2(y), \cdots, x_k = w_k(y)$$

are one-to-one, then the probability distribution of Y is given by

$$g(y) = \sum_{i=1}^{k} f(w_i(y)) \left| \frac{\mathrm{d}w_i}{\mathrm{d}y} \right|$$

- Example: Let $Y = X^2$, then the inverse functions are $w_1 = -\sqrt{y}, w_2 = \sqrt{y}$ that divide the full range where X is defined
  - This gives us the derivatives $-\frac{1}{2\sqrt{y}}, \frac{1}{2\sqrt{y}}$
  - Therefore $g(y) = f(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| + f(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right|$