Supplemental Lecture

4.5 Sum of a random number of random variables

Let $X_1, X_2, \ldots$ be an infinite sequence of independent identically distributed (i.i.d.) random variables. This means each X_i has same PDF (or PMF).

notation: $X_i \sim X \Leftrightarrow X_i, X$ have same PDF (or same PMF)

Let $X \sim X_1, X_2, \ldots\ldots$

Let N be a r.v. with range(N) $= \{0, 1, 2, \ldots\}$. Assume the set $\{N, X_1, X_2, \ldots\}$ is independent. Define

$$Y = X_1 + X_2 + \cdots + X_N$$

Denote by $E[X]$, $Var(X)$ the common mean & variance of X_i.

<u>Goal</u>: find $E[Y]$ and $Var(Y)$

Let $n \geq 0$ be fixed. Then

$$E[Y \mid N=n] = E[X_1 + \cdots + X_N \mid N=n]$$

$$= E[X_1 + \cdots + X_n \mid N = n]$$

$$= E[X_1 + \cdots + X_n] \quad \left\{ \begin{array}{l} \text{by independence} \\ \text{of } N, X_1, X_2, \ldots \end{array} \right.$$

$$= E[X_1] + \cdots + E[X_n]$$

$$= n\, E[X]$$

Thus $E[Y \mid N] = N \cdot E[X]$

By the law of Iterated Expectations

$$E[Y] = E[E[Y \mid N]]$$

$$= E[N \cdot E[X]]$$

$$\therefore \boxed{E[Y] = E[N] \cdot E[X]}$$

Also

$$\mathrm{Var}(Y \mid N=n)$$

$$= \mathrm{Var}(X_1 + \cdots + X_N \mid N=n)$$

$$= \mathrm{Var}(X_1 + \cdots + X_n \mid N=n)$$

$$= \mathrm{Var}(X_1 + \cdots + X_n) \quad \{\text{by indep.}\}$$

$$= \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) \quad \{\text{by indep.}\}$$

$$= n \cdot \mathrm{Var}(X)$$

Thus $\mathrm{Var}(Y \mid N) = N \cdot \mathrm{Var}(X)$.

By the law of total variance

$$\mathrm{Var}(Y)$$

$$= E[\mathrm{Var}(Y|N)] + \mathrm{Var}(E[Y|N])$$

$$= E[N \cdot \mathrm{Var}(X)] + \mathrm{Var}(N \cdot E[X])$$

$$= E[N] \cdot \mathrm{Var}(X) + E[X]^2 \cdot \mathrm{Var}(N)$$

Thus

$$\boxed{\mathrm{Var}(Y) = E[N] \cdot \mathrm{Var}(X) + E[X]^2 \mathrm{Var}(N)}$$

<u>Ex.</u> (lab 4)

Two coins

coin 1 : $P(\text{head}) = p$

coin 2 : $P(\text{head}) = q$

- Flip Coin 1 until 1st head, call # flips N. note $N \sim \text{Geometric}(p)$
- Flip Coin 2 N times, and let Y be the # of heads.

observe

$$Y = X_1 + \cdots + X_N$$

where each $X_i \sim \text{Bernoulli}(q)$, and all indep. let $X \sim X_1$. so we have

$$E[N] = \frac{1}{p}, \quad Var(N) = \frac{1-p}{p^2}$$

and

$$E[X] = q, \quad Var(X) = q(1-q)$$

Therefore

$$E[Y] = E[N] \cdot E[X] = \boxed{\frac{q}{p}}$$

and

$$Var(Y) = E[N] \cdot Var(X) + E[X]^2 \cdot Var(N)$$

$$= \boxed{\frac{q(1-q)}{p} + \frac{q^2(1-p)}{p^2}}$$

$$= \boxed{\frac{q}{p}\left((1-q) + \frac{q}{p}(1-p)\right)}$$

Notation

write $X \sim Y$ to mean $f_X = f_Y$ (or $P_X = P_Y$ in discrete case), equivallently $F_X = F_Y$

Theorem

Let $X, Y \sim \text{Uniform}(0,1)$ be independent. Then

$$\max(X,Y) \sim \sqrt{X}$$

Proof

we are given that

$$f_X(t) = \begin{cases} 1 & \text{if } 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$F_X(t) = \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t > 1 \end{cases}$$

and similarly for Y. Thus, if $0 \le t \le 1$:

$$\begin{aligned} F_{\max(X,Y)}(t) &= P(\max(X,Y) \le t) \\ &= P(X \le t, Y \le t) \\ &= P(X \le t) \cdot P(Y \le t) \quad \Big\{ \text{by indep.} \\ &= F_X(t) \cdot F_Y(t) \\ &= t^2 \end{aligned}$$

Therefore

$$F_{\max(X,Y)}(t) = \begin{cases} 0 & \text{if } t<0 \\ t^2 & 0 \le t \le 1 \\ 1 & t>1 \end{cases}$$

Also

$$F_{\sqrt{X}}(t) = P(\sqrt{X} \le t)$$
$$= P(X \le t^2)$$
$$= F_X(t^2)$$
$$= \begin{cases} 0 & \text{if } t<0 \\ t^2 & \text{if } 0 \le t \le 1 \\ 1 & \text{if } t>1 \end{cases}$$

$\therefore F_{\max(X,Y)} = F_{\sqrt{X}} \quad \therefore \max(X,Y) \sim \sqrt{X}$ ∎

Exercise

- Same hypotheses, show

$$\min(X, Y) \sim 1 - \sqrt{1 - X}$$

$$\nearrow \nearrow$$

$$\sim \text{uniform}(0,1)$$

- Investigate

$$\max(X_1, .., X_n) \sim ?$$

$$\min(X_1, .., X_n) \sim ?$$

where $X_1, .., X_n, X \sim \text{uniform}(0,1)$

- If $X \sim \text{uniform}(a,b)$ then $X + c \sim \text{uniform}(a+c, b+c)$.