Lecture Six of the Stochastic Processes (1) course.

P.G.F. of Sum of Random Number of r.v's

To explain this subject we take this example, if X_i denotes to the number of person's involved the (i^{th}) accident in a day in a certain city, and if the number (N) of accident happening on a day is a random variable, then the sum:

$$S_n = \sum_{i=1}^{n} X_i$$

denoted the total number of persons involved in accidents on a day.

Theorem:

Let X_i, $i = 1, 2, \ldots, n$ be (i.i.d.) random variables with

$$P_r\{X_i = k\} = p_k,$$

and p.g.f. of X_i is:

$$P_{X_i}(S) = \sum_{k=0}^{\infty} p_k S^k$$

and let the sum:

$$S_N = X_1 + X_2 + \cdots + X_N,$$

where N is a random variable independent of the X_i's. Let the distribution of (N) is given by:

$$P_r\{N = n\} = g_n,$$

where the p.g.f. of N is given by:

$$G_N(S) = \sum_{n=0}^{\infty} g_n S^n \qquad \ldots 1.14$$

Then $H(S)$ is the p.g.f. of the sum S_N, and it given by compound function:

$$H(S) = \sum_{j=0}^{\infty} h_j S^j = \sum_{j=0}^{\infty} P_r\{S_N = j\} S^j,$$

where $h_j = P_r\{S_N = j\}$.

Then:

$$H(S) = G_N[P_X(S)] \qquad \ldots 1.15$$

Proof:

Since N is a random variable which can assume values $(0, 1, 2, \ldots)$, then the event: $\{S_N = j\}$ can be happen in the mutually exclusive ways:

$$[N = n \text{ and } S_n = X_1 + X_2 + \cdots + X_n = j], \quad \text{for } n = 1, 2, \ldots$$

Define that $X_0 = S_0 = 0$, so that:

$$X_0 + X_1 + X_2 + \cdots + X_n = S_n, \quad n \geq 0$$

Then:

$$h_j = P_r\{S_N = j\} = \sum_{n=0}^{\infty} P_r\{N = n, S_n = j\}.$$

Since N is independent of X_i's and therefor of S_n, then:

$$h_j = \sum_{n=0}^{\infty} P_r\{N = n\} P_r\{S_n = j\} = \sum_{n=0}^{\infty} g_n P_r\{S_n = j\}.$$

Then the p.g.f. of the sum $S_N = X_1 + X_2 + \cdots + X_N$ is given by:

$$H(S) = \sum_{j=0}^{\infty} h_j S^j = \sum_{j=0}^{\infty} \left[\sum_{n=0}^{\infty} g_n P_r\{S_n = j\} \right] S^j = \sum_{n=0}^{\infty} \left[\sum_{j=0}^{\infty} P_r\{S_n = j\} S^j \right] g_n.$$

Since: $S_n = X_1 + X_2 + \cdots + X_n$ is (i.i.d) r.v's of sum of fixed number n [from theorem (2)], then:

$$H(S) = \sum_{n=0}^{\infty} [P_X(S)]^n g_n = G_N[P_X(S)], \quad \text{compound function.}$$

Note: if $P_X(S)$ and $G_N(S)$ are two p.g.f.'s, then the $G_N[P_X(S)]$ is also p.g.f.

Corollary:

The mean and variance of the sum $S_N = \sum_{i=1}^{N} X_i$, (where N is a random variable) are:

1. $E(S_N) = E(N) \cdot E(X_i)$
2. $\mathrm{Var}(S_N) = E(N) \, \mathrm{Var}(X_i) + [E(X_i)]^2 \, \mathrm{Var}(N)$

Proof:

1. Since $H(S)$ is the p.g.f. of the sum S_N, then:

$$H(S) = G_N(P_X(S)) \quad \text{[theorem (3)].}$$

Then the mean of S_N is:

$$E(S_N) = H'(1) = G'_N(P_X(S)) \cdot P'_X(S)\Big|_{S=1} = G'_N(1) \cdot P'_X(1) \quad \text{since } P_X(1) = 1$$

then

$$E(S_N) = E(N) \cdot E(X_i).$$

2. Since $H(S)$ is the p.g.f. of the sum S_N, then:

$$\text{Var}(S_N) = H''(1) + H'(1) - [H'(1)]^2, \quad \text{(from the definition of variance)}$$

Since:

$$\begin{aligned} H''(1) =& P'_X(S) \cdot G''_N(P_X(S)) \cdot P'_X(S)\big|_{S=1} + G'_N(P_X(S)) \cdot P''_X(S)\big|_{S=1} \\ =& [P'_X(1)]^2 G''_N(1) + G'_N(1) P''_X(1). \end{aligned}$$

Then:

$$\begin{aligned} Var(S_N) =& [P'_X(1)]^2 G''_N(1) + G'_N(1)P''_X(1) + G'_N(1)P'_X(1) - [G'_N(1)P'_X(1)]^2 \\ =& G'_N(1)[P''_X(1) + P'_X(1)] + [P'_X(1)]^2[G''_N(1) - (G'_N(1))^2]. \end{aligned}$$

$$\ldots 1.18$$

By adding and subtracting the amount $[P'_X(1)]^2 G'_N(1)$ to the right side of the equation (1.18), we have:

$$\begin{aligned} Var(S_N) =& G'_N(1)[P''_X(1) + P'_X(1)] - [P'_X(1)]^2 G'_N(1) \\ &+ [P'_X(1)]^2 G'_N(1) + [P'_X(1)]^2[G''_N(1) - (G'_N(1))^2]. \end{aligned}$$

$$Var(S_N) = G'_N(1)[P''_X(1) + P'_X(1) - [P'_X(1)]^2] + [P'_X(1)]^2[G''_N(1) + G'_N(1) - (G'_N(1))^2].$$

$$Var(S_N) = E(N)\,\text{Var}(X_i) + [E(X_i)]^2\,\text{Var}(N),$$

Proof of theorem.

Example:

If $S_N = X_1 + X_2 + \cdots + X_N$ where N has a Poisson distribution with mean λ, and if X_i are (i.i.d) random variables with Geometric distribution:

$$P_r\{X_i = k\} = pq^k,$$

find the p.g.f. of the sum S_N and the mean and variance of it.

Solution:

Since: $N \sim poi(\lambda)$, then the p.g.f. of N is:

$$G_N(S) = e^{\lambda(S-1)}, \quad \text{with mean and variance } \lambda.$$

And since: $X_i \sim Geo(p)$, then the p.g.f. of X_i is:

$$P_{X_i}(S) = \frac{p}{1-qS}, \quad \text{with mean } \frac{q}{p} \text{ and variance } \frac{q}{p^2}.$$

Then the p.g.f. of $S_N = X_1 + X_2 + \cdots + X_N$ is:

$$H(S) = G_N[P_{X_i}(S)] \quad \text{[theorem (3)]}$$

$$= G_N\left[\frac{p}{1-qS}\right] = e^{\lambda\left(\frac{p}{1-qS}-1\right)}, \quad \text{the p.g.f. of } S_N.$$

The mean of S_N is:

$$E(S_N) = E(N)E(X_i) = \lambda\frac{q}{p} = \frac{\lambda q}{p}.$$

The variance of S_N is:

$$\operatorname{Var}(S_N) = E(N)\operatorname{Var}(X_i) + [E(X_i)]^2 \operatorname{Var}(N) = \lambda\frac{q}{p^2} + \left(\frac{q}{p}\right)^2 \lambda = \frac{\lambda q}{p^2}(1+q).$$

Example :

Let $X_1 + X_2 + \cdots + X_N$ be (i.i.d) random variables with Poisson distribution (λ), find the p.g.f. of:

$$S_N = \sum_{i=1}^{N} X_i,$$

where:

1. n is fixed integer.
2. N is random variable and has a Binomial distribution with parameter (n, p).

Solution:

1. Since X_i is Poisson distribution with the same parameter (λ), then the p.g.f. of X_i is:

$$P_{X_i}(S) = e^{\lambda(S-1)}, \quad \text{for all } X_i.$$

Then the p.g.f. of sum i.i.d. (S_N) is:

$$\begin{aligned} P_{S_n}(S) =& [P_X(S)]^n \quad \text{[theorem (2)]} \\ =& [e^{\lambda(S-1)}]^n = e^{n\lambda(S-1)}. \end{aligned}$$

2. Since N is Binomial distribution with parameters (n, p), then the p.g.f. of the random variable N is:

$$G_N(S) = \sum_{n=0}^{\infty} g_n S^n = (pS+q)^n.$$

Then the p.g.f. of (S_N) when N is r.v., is:

$$H(S) = G_N[P_{X_i}(S)] \quad \text{[Theorem]}$$

$$= G_N[e^{\lambda(S-1)}] = (pe^{\lambda(S-1)} + q)^n, \quad \text{the p.g.f. of } S_N.$$

Remark: To find the mean and variance of S_N when N is random variable:

$$E(S_N) = H'(S) = E(N)E(X_i) = np\lambda, \quad \text{the mean of } S_N.$$

The variance of S_N is:

$$\operatorname{Var}(S_N) = E(N)\operatorname{Var}(X_i) + [E(X_i)]^2 \operatorname{Var}(N) = np\lambda + \lambda^2 npq = \lambda np(1 + \lambda q).$$

H.W: From example (1.14), find the mean and variance of S_N when:

$$X_i \sim Poi(2) \text{ and } N \sim Bin(30, \tfrac{1}{2}).$$