Probability Generating Functions

Suppose X is a random non-negative variable

$G_x(t) = E[t^x]$ $G_x(t)$ is a probability generating function (PGF)

the r.v.

t is a dummy variable (you'll see) → acting in place of some real value.

recall $E[g(x)] = \sum_x g(x) P(X=x)$

$\Downarrow$

$$G_x(t) = E[t^x] = \sum_x t^x P(X=x) = \underbrace{P(X=0) + P(X=1)t + P(X=2)t^2 + \dots}_{\text{polynomial in } t}$$

$G_x(0) = P(X=0)$ $G_x(1) = P(X=0) + P(X=1) + P(X=2) + \dots = 1$

$$G_x'(t) = \frac{d}{dt} G_x(t) = \frac{d}{dt}\sum_{x=0}^{\infty} t^x P(X=x) = \sum_{x=0}^{\infty} P(X=x) \cdot \frac{d}{dt} t^x = \sum_{x=0}^{\infty} x t^{x-1} P(X=x)$$

$= P(X=1) + 2P(X=2)t + \dots$ (kind of like MacLaurin series...)

$$G_x'(1) = \sum_{x=0}^{\infty} x(1)^{x-1} P(X=x) = \sum_{x=0}^{\infty} x P(X=x) = E[X]$$

⚠ $G_x(0) = P(X=0)$

$G_x(1) = 1$

$G_x'(1) = E[X]$

$P(X=k) = \dfrac{G_x^{(k)}(0)}{k!}$

differentiate and sub $t=0$

to get coefficients/probabilities

(not in formula book ☹)

$X \sim B(n,p)$ $\quad P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

$$G_X(t) = E[t^X] = \sum_{x=0}^{n} t^x P(X=x) = \sum_{x=0}^{n} t^x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pt)^x (1-p)^{n-x}$$

comparing with

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} \Rightarrow \underbrace{G_X(t) = (pt+1-p)^n}_{\text{PGF for Binomial Dist}^{\underline{n}}}$$

All the PGF's are in the formula book! (Page 23)

$X \sim Po(\lambda)$ $\quad P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\quad x = 0,1,2,3\ldots$

(maclaurin series)

$$G_X(t) = \sum_{x=0}^{\infty} t^x e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(\lambda t)^x}{x!} \quad \xleftrightarrow[\text{with}]{\text{comparing}} \quad \sum_{k=0}^{\infty}\frac{y^k}{k!} = e^y$$

$\Downarrow$

$$G_X(t) = e^{-\lambda}e^{\lambda t}$$

$$G_X(t) = e^{\lambda(t-1)} \leftarrow \text{PGF for Poisson Dist}^{\underline{n}}$$

$X \sim Geo(p)$ $\quad P(X=x) = p(1-p)^{x-1}$ $\quad x = 1,2,3,4,5\ldots$

$$G_X(t) = \sum_{x=1}^{\infty} t^x p(1-p)^{x-1} = p\sum_{x=1}^{\infty} t^x (1-p)^{x-1} = \frac{p}{1-p}\sum_{x=1}^{\infty}[t(1-p)]^x$$

comparing with

$S_\infty = \frac{a}{1-r}$, $|r|<1$

geometric series

$$G_X(t) = \frac{p}{1-p}\,\frac{t(1-p)}{1-t(1-p)}$$

$$\underbrace{G_X(t) = \frac{pt}{1-t+pt}}_{\text{PGF for Geometric dist}^{\underline{n}}}$$

$X \sim NB(r,p)$ $\quad P(X=x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}$ $\quad x = r, r+1, r+2\ldots$ $\quad$ let $q = 1-p$ (prob. of failure)

$$G_X(t) = \sum_{x=r}^{\infty} t^x \binom{x-1}{r-1} p^r (1-p)^{x-r} = p^r \sum_{x=r}^{\infty}\binom{x-1}{r-1} t^{x-r}(1-p)^{x-r} t^r = (pt)^r \sum_{x=r}^{\infty}\binom{x-1}{r-1}(t-tp)^{x-r}$$

compare with

$$G_X(t) = (pt)^r(1-[t(1-p)])^{-r}$$

$\sum_{x=r}^{\infty}\binom{x-1}{r-1}p^r q^{x-r} = 1$

Sum of probabilities $=1$

$$\sum_{x=r}^{\infty}\binom{x-1}{r-1}q^{x-r} = p^{-r} = (1-p)^{-r}$$

PGF for Negative Binomial dist$^{\underline{n}}$

$$\left\{\, G_X(t) = \left(\frac{pt}{1-t+pt}\right)^r\right.$$

Ex 7A

1. a) $G_X(t) = 0.3 + 0.2t + 0.5t^2$

x	0	1	2
$p(X=x)$	0.3	0.2	0.5

$\sum_{x=0}^{r} t^x p(X=x)$

Sample space = $\{0, 1, 2\}$

b) $p(X=0) = 0.3$ $\quad p(X \geq 0) = 1$

3. a) $G_Y(t) = 0.7 + 0.1(t^2 + t^3 + t^5)$

y	0	1	2	3	4	5
$p(Y=y)$	0.7	0	0.1	0.1	0	0.1

$p(Y=1) = 0$

$p(Y<3) = p(Y=0) + p(Y=2) = 0.7 + 0.1 = 0.8$

$p(3 \leq Y \leq 6) = 0.1 + 0.1 = 0.2$

6. a) $p(X=x) = \frac{x}{10}$ $\quad G_X(t) = \sum_{x=0}^{\infty} t^x p(X=x) = 0 + \frac{1}{10}t + \frac{2}{10}t^2 + \frac{3}{10}t^3 + \frac{4}{10}t^4 = \frac{1}{10}t + \frac{1}{5}t^2 + \frac{3}{10}t^3 + \frac{2}{5}t^4$

10. $G_X(t) = 0.1(2t + 5t^2 + 4t^3) \longrightarrow$ Adds to >1 (1.1), <u>NOT</u> a PGF

11. $G_Y(t) = k(1+t)^{10}$

a) $G_Y(1) = k(2)^{10} = 1$

$k = \frac{1}{1024}$

b) Largest val. of Y $\quad G_Y(t) = \frac{1}{1024}\left(1 + 10t + \ldots + \binom{10}{10}t^{10}\right) = \frac{1}{1024} + \frac{10}{1024} + \ldots + \frac{1}{1024}t^{10}$

$Y_{max} = 10 \quad p(Y=10) = \frac{1}{1024}$

c) $P(Y=5) = \frac{1}{1024}\binom{10}{5} = \frac{63}{256} = 0.246$

d) $P(Y=y) = \binom{10}{y}\frac{1}{1024} = \binom{10}{y}\left(\frac{1}{2}\right)^{10} = \binom{10}{y}\left(\frac{1}{2}\right)^{y}\left(1 - \frac{1}{2}\right)^{10-y} = \binom{n}{x} p^x (1-p)^{n-x}$

$\therefore Y \sim B(10, 0.5)$

$$G_X(t)=\sum_{x=0}^{\infty} t^x P(X=x)$$

How do we get variance?

$$G_X'(t)=\sum_{x=0}^{\infty}\frac{d}{dt}t^x P(X=x)=\sum_{x=0}^{\infty} x t^{x-1}P(X=x) \Rightarrow G'(1)=\sum_{x=0}^{\infty} xP(X=x)=E[X]$$

$$G_X''(t)=\sum_{x=0}^{\infty}\frac{d}{dt}x t^{x-1}P(X=x)=\sum_{x=0}^{\infty} x(x-1)t^{x-2}P(X=x)$$

$$=\sum_{x=0}^{\infty} x^2 t^{x-2}P(X=x)-\sum_{x=0}^{\infty} x t^{x-2}P(X=x)$$

$$\Rightarrow G_X''(1)=\sum_{x=0}^{\infty} x^2 P(X=x)-\sum_{x=0}^{\infty} xP(X=x)=E[X^2]-E[X]$$

almost there!

$$Var(X)=E[X^2]-(E[X])^2=E[X^2]-E[X]+E[X]-(E[X])^2$$

$$\boxed{=G_X''(1)+G_X'(1)-[G_X'(1)]^2}$$

Suppose $Z=X+Y$ and X independent of Y $(X\perp Y)$

$$G_Z(t)=E[t^Z]=E[t^{X+Y}]=\sum_x\sum_Y t^{x+y}P(X+Y=x+y)=\sum_x\sum_Y t^x t^y P(X=x)P(Y=y)$$

$$=\sum_x t^x P(X=x)\sum t^y P(Y=y)=G_X(t)G_Y(t)$$

(Proof is out of syllabus)

→ To generalise, if $Y=X_1+X_2+X_3+\dots+X_n$ $\quad X_i\perp X_j,\ i\neq j$

$$\Rightarrow G_Y(t)=G_{X_1}(t)\cdot G_{X_2}(t)\cdot G_{X_3}(t)\cdot\ \dots\ \cdot G_{X_n}(t)=\prod_{i=1}^{n}G_{X_i}(t)$$

this you need to know

Sum of independent distributions has PGF of products of PGF's (if that makes sense)

If X has PGF $G_X(t)$ and $Y=aX+b$,

$$G_Y(t)=E[t^Y]=\sum_Y t^y P(Y=y)=\sum_x t^{ax+b}P(aX+b=y)=\sum_x t^b(t^a)^x P\left(X=\frac{y-b}{a}\right)$$

$$=t^b\sum_x (t^a)^x P(X=x)=t^b G_X(t^a)$$ ← PGF of linear transformation of X

Ex (7D)

7. Aidan & Chloe each have 5 scratchcards

Aidan : $P(\text{Win}) = 0.3$ Chloe : $P(\text{Win} = 0.4)$

$X =$ total # of scratchcards won

a) $A =$ # of wins by Aidan $A \sim Bin(5, 0.3)$ $G_A(t) = (0.7 + 0.3t)^5$

$C =$ " " " " Chloe $C \sim Bin(5, 0.4)$ $G_C(t) = (0.6 + 0.4t)^5$

$X = A + C$ $\therefore G_X(t) = G_A(t) + G_C(t) = (0.7 + 0.3t)^5 + (0.6 + 0.4t)^5$

b) $G_X'(t) = 5 \times 0.3 \times (0.7 + 0.3t)^4 + 5 \times 0.4 \times (0.6 + 0.4t)^4 = 1.5(0.7 + 0.3t)^4 + 2(0.6 + 0.4t)^4$

$\bar{X} = G_X'(1) = 1.5 + 2 = 3.5$