Poisson Distribution

\# of babies born in 24 hours
\# of chocolate chips in a cookie
\# of pepperoni slices on a pizza

X = # of events in an interval

eg number of chocolate chips

eg in the volume of a cookie

Assumptions for a Poisson Distribution

Events occur: independently: event occuring doesn't affect probability

singly: events occur one at a time

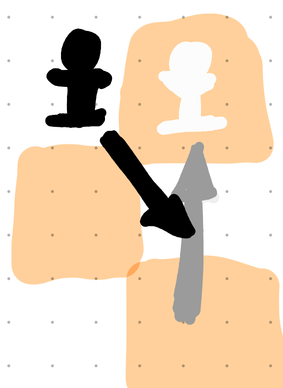
at a constant average rate: # of events $\propto$ interval size

X = # of events in an interval $X \sim Po(\lambda)$ $P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

X can assume x where x = 0, 1, 2 ... (no upper bound)

Special property: $E[X] = Var(X) = \lambda$

Proof on next page



En Poisson.

Proof that $E[X]=\lambda$ for $X\sim Po(\lambda)$

$$e^x=\sum_{r=0}^{\infty}\frac{x^r}{r!}=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\dots$$

when $x=0$, $\frac{xe^{-\lambda}\lambda^x}{x!}=0$, so we can start at $x=1$

$$E[X]=\sum_{x=0}^{\infty}xP(X=x)=\sum_{x=0}^{\infty}\frac{xe^{-\lambda}\lambda^x}{x!}=\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^x}{(x-1)!}=e^{-\lambda}\sum_{x=1}^{\infty}\frac{(\lambda)(\lambda^{x-1})}{(x-1)!}$$

$$=\lambda e^{-\lambda}\sum_{t=0}^{\infty}\frac{\lambda^t}{t!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

redifine $t=x-1$ so we start at 0 and can use Maclaurin

Proof that $Var(x)=\lambda$ for $X\sim Po(\lambda)$

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

Consider $E[X(X-1)]=E[X^2-X]=E[X^2]-E[X]$

$$\Rightarrow Var(x)=E[X(X-1)]+E[X]-(E[X])^2$$

$$E[X(X-1)]=\sum_{x=0}^{\infty}x(x-1)\frac{e^{-\lambda}\lambda^x}{x!}=\underbrace{e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!}=\lambda^2e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}=\lambda^2e^{-\lambda}\sum_{t=0}^{\infty}\frac{\lambda^t}{t!}}_{\text{Do the same stuff as proof of } E[X]=\lambda}$$

$$=\lambda^2e^{-\lambda}e^{\lambda}=\lambda^2$$

$E[X(X-1)]=\lambda^2$, $E[X]=\lambda$ $\Rightarrow$ $Var(x)=\lambda^2+\lambda-\lambda^2=\lambda$

Ex 2C

10 Flaw in a material: 1.5 per 50m

a) X = # of flaws per 50m $\quad X \sim Po(1.5)$

$$P(X=3) = \frac{e^{-1.5}(1.5)^3}{3!} = 0.1255$$

b) Material sold in rolls of 200m

X' = # of flaws per 200m $\quad X' \sim Po(6)$

$P(X'<4) = P(X' \leq 3) = 0.1512$

c) 5 rolls of meterial

Y = # of rolls with <4 flaws out of 5 rolls

$Y \sim B(5, 0.1512) \qquad P(Y \geq 2) = 1 - P(Y \leq 1) = 0.8330$

Define a new variable with a binomial distribution with $p=0.1512$ to find probability of at least 2 rolls out of 5 have less than 4 flaws.

14 Machine: 1.5 breakdowns / week

a) X = # of breakdowns per week

$X \sim Po(1.5) \qquad P(X \leq 2) = 0.8088$

b) X' = # of breakdowns per 2 weeks

$X' \sim Po(3) \qquad P(X' \geq 5) = 1 - P(X' \leq 4) = 1 - 0.8153 = 0.1847$

c) A firm will repair breakdowns of 6-weeks.

If more than n breakdowns / 6 weeks, full refund.

firm wants probability of full refund $< 5\%$ or 0.05

X'' = # of breakdowns / 6-weeks $\quad X'' \sim Po(9)$

$P(\text{refund}) = P(X'' > n) = 1 - P(X'' \leq n) < 0.05$

$P(X'' \leq n) > 0.95$

Smallest n to satisfy is 14 $\quad (P(X'' \leq 14) = 0.9585)$

Sum of Individual Poisson r.v.'s

let $X \sim Po(\lambda)$, $Y \sim Po(\mu)$ $\quad X \perp Y$ (X and Y are independent)

$x = 0, 1, 2 \ldots$ $\quad y = 0, 1, 2 \ldots$ $\quad$ let $Z = X + Y$

$$P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad P(Y=y) = \frac{e^{-\mu}\mu^{y}}{y!}$$

$$P(Z=z) = \sum_{x=0}^{z} P(X=x \cap Y=z-x)$$

for $Z=z$, $x+y=z$ for each value of x, $y=z-x$

since $X \perp Y$

$$= \sum_{x=0}^{z}\left(P(X=x) \times P(Y=z-x)\right) = \sum_{x=0}^{z}\left(\frac{e^{-\lambda}\lambda^{x}}{x!} \cdot \frac{e^{-\mu}\mu^{z-x}}{(z-x)!}\right)$$

$$= \frac{e^{-(\lambda+\mu)}}{z!}\sum_{x=0}^{z}\left(\frac{z!\,\lambda^{x}\mu^{z-x}}{x!\,(z-x)!}\right)$$

$\left(\frac{z!}{z!} \text{ multiplied}\right)$

$$= \frac{e^{-(\lambda+\mu)}}{z!}\sum_{x=0}^{z}\left({}^{z}C_{x} \cdot \lambda^{x}\mu^{z-x}\right) = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^{z}}{z!} \Rightarrow Z \sim Po(\lambda+\mu)$$

$${}^{z}C_{x} = \frac{z!}{x!\,(z-x)!}$$

$$\sum_{x=0}^{z} {}^{z}C_{x} \cdot \lambda^{x}\mu^{z-x}$$

$= (\lambda+\mu)^{z}$ (binomial expansion)

Therefore,

$X + Y \sim Po(\lambda+\mu)$ (if $X \perp Y$)

This derivation is not needed in the exam, but is fun.

Ex (2D)

(1) $X \sim Po(3.3)$ $Y \sim Po(2.7)$ $X+Y \sim Po(6)$

a) $P(X+Y=5) = 0.1606$

b) $P(X+Y \leq 7) = 0.7440$

c) $P(X+Y>4) = 1-P(X+Y \leq 4)$

$= 1-0.2851 = 0.7149$

(3) $X \sim Po(2.5)$ $Y \sim Po(3.5)$

a) $P(X=2 \cap Y=2) = 0.2565 \times 0.1850 = 0.0475$

b) $P(X>2 \cap Y>2) = (1-P(X \leq 2))(1-P(Y \leq 2)) \neq P(X+Y>4)$

$= (1-0.5438)(1-0.3208) = (0.4562)(0.6792) = 0.3099$

(6) $A \sim Po(1.25)$ $B \sim Po(0.75)$ $A+B \sim Po(2)$

a) $P(A=2) = 0.2238$ b) $P(A+B=2) = 0.2707$

c) Total taxis in 5 days

$= A' + B' \sim Po(5 \times 1.25 + 5 \times 0.75 = 10)$

$P(A'+B'<10) = P(A'+B' \leq 9) = 0.4579$

(10) P = # of personal emails / hour $P \sim Po(1.8)$

B = # of business emails / hour $B \sim Po(3.7)$

A = # of advertising emails / hour $A \sim Po(1.5)$

a) $P' \sim Po(0.9)$ $B' \sim Po(1.85)$ $A' \sim Po(0.75)$

$P(P' \geq 1 \cap B' \geq 1 \cap A' \geq 1) = (1-P(P'=0))(1-P(B'=0))(1-P(A'=0))$

$= (1-0.4066)(1-0.1572)(1-0.4724) = 0.5934 \times 0.8428 \times 0.5276 = 0.2639$

b) $P'' \sim Po(14.4)$ $B'' \sim Po(29.6)$ $A'' \sim Po(12)$ $P''+B''+A'' \sim Po(56)$ (8 hours)

$P(P''+B''+A''>50) = 1-P(P''+B''+A'' \leq 50) = 1-0.2343 = 0.7657$

c) X = # of days she recieves > 50 emails out of 5 days

$X \sim B(5, 0.7657)$ $P(X=2) = 0.07541$ BINOMIAL MODEL

How do you know if $X \sim Po(\lambda)$?

How do you know if $X \sim Po(\lambda)$ is a suitable model?

$\downarrow$

$E[X] = Var(X) = \lambda$ ← if satisfied, $X \sim Po(\lambda)$ or can be modelled accurately using $Po(\lambda)$

(just a sanity check)

Variance reminder:

$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$ (if list of data)

$= \frac{\sum fx^2}{\sum f} - \bar{x}^2$ (if given as frequency table)

= mean of squares − square of mean

Ex 2E

3 a) $\bar{x} = \frac{\sum fx}{\sum f} = \frac{0+19+54+75+76+55+42+21}{8+19+28+25+19+11+7+3} = 2.8\dot{6}\dot{6}$ flaws per piece

$\sigma^2 = \frac{\sum fx^2}{\sum f} - 2.85^2 = \frac{19+4\times28+9\times25+16\times19+25\times11+36\times7+49\times3}{120} - 2.85^2$

$= 11.116\dot{6} - 2.86\dot{6}^2 = 2.899$

b) $\bar{x} \approx \sigma^2$ $\quad 2.866 \approx 2.899$

c) in sample of 120 cloths, frequency of ≥ 8 is 0, but it doesn't mean impossible.

d) let $X \sim Po(2.899)$ where X = # of flaws per cloth

# of cloth with ≥ 8 flaws in 10000 pieces of cloth

$= 10000 \times P(X \geq 8) = 10000 - 10000 \times P(X \leq 7) = 10000 - 10000(0.99013431\ldots)$

$= 10000 - 9901.3431 = 98.656896\ldots \approx 99$ pieces with ≥ 8 flaws

Back to binomial distribution for a sec...

$X \sim B(n, p)$

Epected value of binomial distribution.

$$E[X] = \sum_{x=0}^{n} x P(X=x) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{x=1}^{n} {}^{n-1}C_{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{y=0}^{n-1} {}^{n-1}C_{y} p^{y} (1-p)^{(n-1)-y}$$

$$= np \underbrace{(p+(1-p))}_{=1}{}^{n-1} = np$$

if $X \sim B(n,p)$,

$E[X] = np$

Variance of binomial distribution

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

$$E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X]$$

Then, $Var(X) = E[X(X-1)] + E[X] - (E[X])^2$

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= n(n-1)p^2 \sum_{x=0}^{n} \frac{(n-2)!}{(x-2)!((n-2)-(x-2))!} p^{x-2} (1-p)^{(n-2)-(x-2)}$$

$$= n(n-1)p^2 \sum_{t=0}^{n-2} {}^{n-2}C_t \, p^t (1-p)^{(n-2)-t}$$

$$= n(n-1)p^2 (p + (1-p))^{n-2} = n(n-1)p^2$$

$$Var(X) = n^2p^2 - np^2 + np - n^2p^2 = np - np^2 = np(1-p)$$

if $X \sim B(n,p)$, $Var(X) = np(1-p)$

Why did we go back to binomial distribution?

approximating the binomial distribution using Poisson distribution.

If $X \sim B(n,p)$, n is large and p is small,

X can be approximated with $Po(np)$

For approximation, $E[X] \approx Var(x)$

$np \approx np(1-p)$

p is small, $1-p \approx 1$, $np \approx np(1-p)$

n is large, $np \approx np(1-p)$

How large/small is large/small enough?

¯_(ツ)_/¯ the question will hint towards an approximation

"$p \to 0$ and $n \to \infty$ such that np remains roughly constant"

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X=x) = \lim_{\substack{n\to\infty \\ p\to 0}} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}, \text{ where } \lambda = np$$

$$= \lim_{n\to\infty} \frac{\lambda^x}{x!} \frac{n!}{(n-x)!\,n^x} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{x!} \underbrace{\left(\lim_{n\to\infty} \frac{n!}{(n-x)!\,n^x}\right)}_{\text{tends to what?}} \underbrace{\left(\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n\right)}_{\text{tends to } e^{-\lambda}} \underbrace{\left(\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-x}\right)}_{\text{tends to } 1^{-x} = 1}$$

$$= \frac{\lambda^x}{x!} e^{-\lambda} \underbrace{\left(\lim_{n\to\infty} \frac{n!}{(n-x)!\,n^x}\right)}_{\text{need to show it tends to 1}}$$

$$= \left.\frac{\lambda^x e^{-\lambda}}{x!}\right\} \text{Poisson distribution.}$$

$$\boxed{\begin{aligned} \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n &= \lim_{n\to\infty}\left(1+x+\frac{n(n+1)x^2}{2!\,n^2}+\frac{n(n+1)(n+2)x^3}{3!\,n^3}+\dots\right) \\ &= 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\dots = e^x \\ &\left(\lim_{n\to\infty}\left(n(n+1)(n+2)\right) = n^3\right) \end{aligned}}$$

Consider $$\lim_{n\to\infty} \frac{n!}{(n-x)!\,n^x} = \lim_{n\to\infty} \frac{\overbrace{n(n-1)\dots(n-(x-1))}^{\text{tends to } n^x \text{ when } n\to\infty}\cancel{(n-x)!}}{\cancel{(n-x)!}\; n^x}$$

$$= \frac{n^x}{n^x} = 1$$

Ex 2G

3. $Y \sim B(200, 0.98)$ $X = 200 - Y$ $X \sim B(200, 0.02)$

b) $X \stackrel{\cdot}{\sim} Po(200 \times 0.02 = 4)$ $\stackrel{\cdot}{\sim}$ means approximately distributed as

$P(Y=197) \approx P(X=3) = \dfrac{e^{-4}\,4^3}{3!} = 0.195$

$P(Y \geq 198) \approx P(X \leq 2) = 0.2381$