Moment Generating Function of $X : S \longrightarrow \mathbb{R}$.

$$\psi_X(t) = E(e^{tX})$$

$$= E\left(\sum_{j=0}^{\infty} (tX)^j / j!\right)$$

$$= \sum_{j=0}^{\infty} t^j E(X^j) / j!$$

$$= \begin{cases} \sum_r f_X(r) e^{tr}, & \text{discrete} \\ \int_{-\infty}^{\infty} f_X(r) e^{tr} \, dr, & \text{cont. case} \end{cases}$$

Thm If $X, Y: S \to \mathbb{R}$
and $\psi_X(t) = \psi_Y(t)$ for all t
in an open interval, then $f_X = f_Y$.

Thm If $X_1, \dots, X_n : S \to \mathbb{R}$ are
independent, $\psi_{X_1+\dots+X_n}(t) = \prod_i \psi_{X_i}(t)$

Why?

$$\psi_{X_1+\dots+X_n}(t) = E\left(e^{t(X_1+\dots+X_n)}\right) = E\left(e^{tX_1} \cdots e^{tX_n}\right) = E\left(e^{tX_1}\right) \cdots E\left(e^{tX_n}\right) = \prod_i \psi_{X_i}(t)$$

$X_1, \dots, X_n$ independent

Example: $X : S \to \mathbb{R}$ Poisson

$$f_X(j) = e^{-\lambda} \lambda^j / j!$$

for $j = 0, 1, \dots$

$$\psi_X(t) = E(e^{tX}) = \sum_{j=0}^{\infty} f_X(j)\, e^{tj}$$

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

$$= e^{-\lambda} \sum_{j=0}^{\infty} (\lambda e^t)^j / j!$$

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

<u>Cor</u>: If $X_1, \ldots, X_\ell : S \to R$ independent Poisson rates $\lambda_1, \ldots, \lambda_\ell$

Then $X = X_1 + \cdots + X_\ell$ = Poisson, rate $\lambda = \lambda_1 + \cdots + \lambda_\ell$

<u>Why</u>: $\psi_X(t) = \prod \psi_{X_i}(t)$

$$= \prod_i e^{\lambda_i(e^t - 1)} = e^{(\sum_i \lambda_i)(e^t - 1)}$$

Example: Stifling Outbreaks

Suppose ℓ people in one location have covid-19

$\lambda_0 = \lambda(X_i)$ = rate at which i^{th} person creates new cases over 2 weeks

X_i = # cases created by i^{th} person

$X = X_1 + \cdots + X_\ell$ = # cases resulting from group

Given $0 < \varepsilon < 1$ how small must λ_0 be for $P(X=0) \geq \varepsilon$?

X Poisson, $\lambda_X = \ell \lambda_0$

$$P(X=0) = e^{-\lambda(X)} \frac{\lambda(X)^0}{0!} = e^{-\ell \lambda_0}$$

Need: $e^{-\ell\lambda_0} \geq \varepsilon$

$$-\ell \lambda_0 \geq \log \varepsilon, \quad \lambda_0 \leq \frac{-\log \varepsilon}{\ell}$$

$\ell = 10$, $\varepsilon = 1/2$, $\lambda_0 \leq \frac{\log 2}{10} = 0.069$

Covariance $X, Y : S \to \mathbb{R}$

$$Cov(X,Y) = E\big((X - E(X))(Y - E(Y))\big)$$
$$= E(XY) - E(X)E(Y)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$$

Significance $Cov(X,Y)$ large if $X - E(X)$ and $Y - E(Y)$ have same sign when large.

Correlation $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sigma_X \, \sigma_Y}$

Fact: $-1 \leq \rho(X,Y) \leq 1$.

de Groot

Ex: $Y = aX + b$, a, b constant, $a \neq 0$

$$\rho(X,Y) = \begin{cases} 1 & \text{if } a > 0 \\ -1 & \text{if } a < 0 \end{cases}$$

<u>Proof</u> $Cov(X, Y) = E((X-E(X))(Y-E(Y)))$

$$Y = aX + b \qquad E(Y) = aE(X) + b$$

$$Y - E(Y) = a(X - E(X))$$

$$Cov(X, Y) = E(a(X - E(X))^2) = a\,Var(X)$$

$$\rho(X, Y) = Cov(X, Y) / \sigma_X \sigma_Y$$

$$\sigma_X = \sqrt{Var(X)} \geq 0$$

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{E((Y - E(Y))^2)}$$

$$= \sqrt{a^2 E((X - E(X))^2)}$$

$$= \sqrt{a^2 Var(X)}$$

$$= |a| \cdot \sigma_X$$

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{a\,Var(X)}{|a|\,\sigma_X^2} = \frac{a}{|a|}$$

$$= \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

Warning: $\rho(X, Y)$ measures **linear** relationships between X and Y, not all possible dependencies.

Ex: X takes values $\{-1, 0, 1\}$ with probability $1/3$ each

$Y = X^2$ not independent!

$$f_{X,Y} \neq f_X \cdot f_Y$$

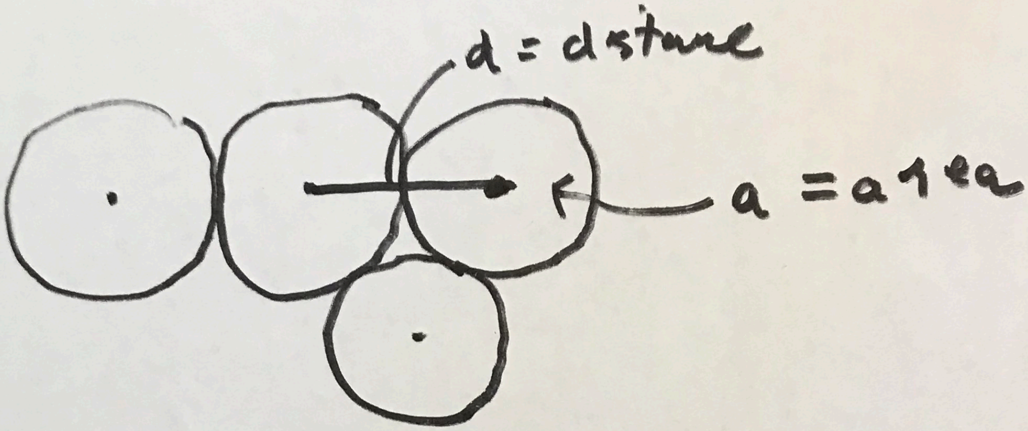
$$\begin{aligned} \mathrm{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(X^3) - E(X)E(Y) \\ &= E(X) - E(X)E(Y) = 0 \end{aligned}$$

$\rho(X, Y) = 0$

Ex: S = set of US cities

$X : S \to \mathbb{R}$ prop. of pop. with covid-19

$Z : S \to \mathbb{R}$ pop. density, avg. square feet per person.



<u>Ask</u>: Does prop. of pop. with covid-19 vary linearly with average distance or average area per person?

Distance: $\rho(X, \sqrt{Z})$

Area: $\rho(X, Z)$

Log normal Distributions + Log Central Limit Theorem

$Y_1, \dots, Y_n$ independent, same density function

$$Y = Y_1 \cdot Y_2 \cdots Y_n$$

$$\log Y = \log Y_1 + \dots + \log Y_n$$

<u>Central Limit Thm</u>: As $n \to \infty$

$\log Y$ has density function tending to a normal distribution X with

$$E(X) = n\, E(\log Y_i) \quad \text{any } i$$

$$\sigma_X^2 = \mathrm{Var}(X) = n\, \mathrm{Var}(\log Y_i)$$

<u>Def</u>. Y tends to log normal random variable $Z = e^X$, $\log f_Z = f_X$ = normal

<u>Fact</u>:

$$E(Y) = e^{E(X) + \sigma_X^2/2}$$

$$Var(Y) = e^{2E(X) + \sigma_X^2}(e^{\sigma_X^2} - 1)$$

Scenario: At some starting time, say we have an amount A in a bank. Flip a fair coin n times. If get a head on a flip, multiply what you have by $r > 1$, if get a tail mult. by $r^{-1} < 1$.

After n days what is the expect amount in the account?

$$S = \{\Delta = (\Delta_1, \ldots, \Delta_n) ; \Delta_i \in \{-1, 1\}\}$$

$$Y_i : S \to \mathbb{R}, \quad Y_i(\Delta) = r^{\Delta_i}$$

$$Y = Y_1 \cdots Y_n \text{ , final amount } YA$$

$$E(YA) = A\,E(Y) \cong$$

$$\cong A\, e^{E(X) + \sigma_X^2/2}$$

$$X = \log Y_1 + \dots + \log Y_n$$

$$E(X) = n\,E(\log Y_i) = n\left(\tfrac{1}{2}\log r + \tfrac{1}{2}\log r^{-1}\right)$$

$$= 0$$

$$\sigma_X^2 = \mathrm{Var}(X) = n\,\mathrm{Var}(\log Y_i)$$

$$= n\left(E\left((\log Y_i)^2\right)\right)$$

$$= n \cdot \left(\tfrac{1}{2}(\log r)^2 + \tfrac{1}{2}(-\log r)^2\right) = n(\log r)^2$$

$$E(Y) = e^{E(X) + \sigma_X^2/2} = e^{n(\log r)^2/2} > 1$$

<u>Note</u>: If $1 < r < e$, $0 < \log r < 1$,

$$(\log r)^2 < \log r \text{ ,}$$

$$e^{n(\log r)^2/2} < e^{n \log r/2} = r^{n/2}$$

<u>Conclusion</u>: If we use coin flips with a fair coin to go up or down by a factor of r or r^{-1}, we'll end up increasing the account after n days!

A coin flip $\Delta = (\Delta_1, \ldots, \Delta_n)$ might have same # of heads + tails.

Outcome in this case $r^{\#\text{heads}}\, r^{-(\#\text{tails})} = 1$

<u>But</u> if Δ has a more heads than tails there's a twin Δ' coming from switching heads + tails.

Δ gives factor r^{a}

Δ' gives factor r^{-a}

$r^{a} + r^{-a} > 2$ if $r > 1, a > 0$