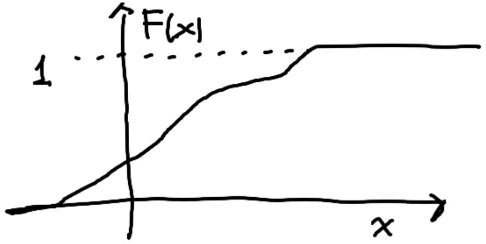
# Seminar 1

- $X, x$ – rand. var, $X$ takes $x$ value : $X = x$
- $\Omega$ set of possible values
- $p(x)$ prob density function, $p(x) \geq 0, \forall x \in \Omega$
- CDF or $F(x)$ cumulative distribution func

$$CDF(x) = \int_{-\infty}^{x} p(x')\, dx'$$

1. $F(x) : \mathbb{R} \to [0, 1]$
2. $F(x_1) \leq F(x_2)$ if $x_1 < x_2$




- Moments : $\mathbb{E}[f(x)] = \langle f(x) \rangle = \int_{\Omega} f(x) \cdot p(x)\, dx'$ – general rule

n-th moment $\langle x^n \rangle = \int_{\Omega} x^n p(x)\, dx = \mu_n$

mean $\langle x \rangle = \mu_1$ variance $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \mu_2 - \mu_1^2$ $\sigma$ - std

# Inequalities

## Markov

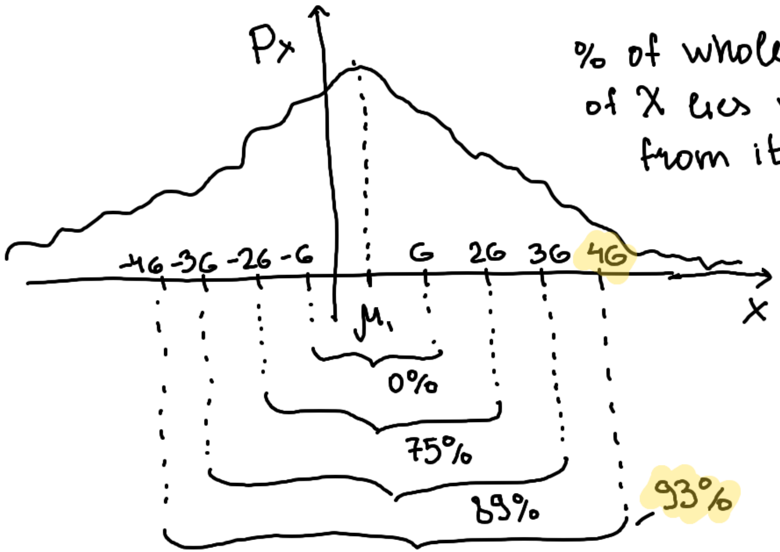
for non-negative R.V $X > 0$, $c > 0$

$$\Pr\{X \geq c\} \leq \frac{\mathbb{E}[x]}{c} \quad \text{OR} \quad \Pr\{X \geq k \cdot \langle X \rangle\} \leq \frac{1}{k}$$

## Chebyshev

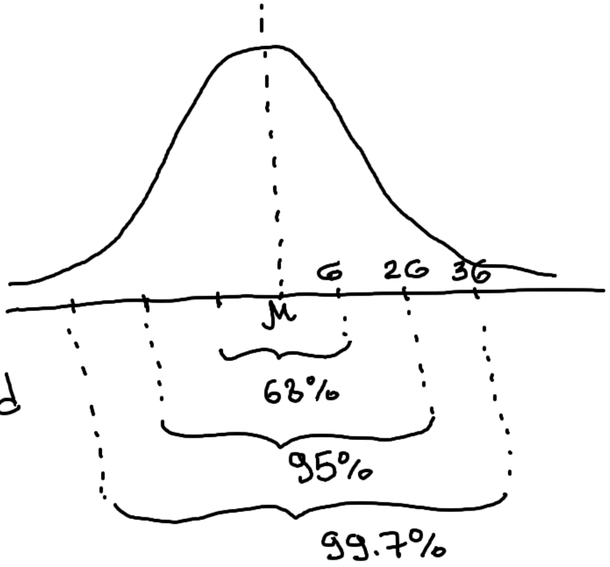
for any distribution
with finite $\mu_1$ and $\mu_2$

$$\Pr\{|X - \langle X \rangle| \geq c\} \leq \frac{\sigma^2}{c^2} \quad \text{OR} \quad \Pr\{|X - \langle X \rangle| \geq k \cdot \sigma\} \leq \frac{1}{k^2}$$




% of whole population
of $X$ lies within $k\sigma$
from its mean

for Gaussian
it can be improved
(68-95-99 RULE OR
3σ RULE)



## Characteristic function

$$\mathcal{G}(k) = \text{F.T.}[p(x)]$$

$$= \int_\Omega e^{ikx} p(x)\, dx$$

omitting $i$ gives moment-generating func (Laplas transform)

$$= \langle e^{ikx} \rangle_x$$

$$\begin{tikzcd}
\text{pdf } p(x) \arrow[r, "FT"] \arrow[d, "\langle \dots \rangle_x"'] & \mathcal{G}(k) \arrow[dl, "\partial^n .. |_{k=0}"] \\
\mu_n &
\end{tikzcd}$$

① if $X_1 \dots X_n$ - independent RV not necessarily identically distributed

then $S_n = \sum_{i=1}^{n} a_i X_i$ has $\mathcal{G}_S(k) = \mathcal{G}_{X_1}(a_1 k) \cdot \dots \cdot \mathcal{G}_{X_n}(a_n k)$

② Relation to moments :
$$\frac{d^n}{dk^n} \mathcal{G}(k) = \frac{d^n}{dk^n} \langle e^{ikx} \rangle_x = \langle (ix)^n e^{ikx} \rangle_x$$

$$\left.\frac{d^n}{dk^n} \mathcal{G}(k)\right|_{k=0} = \langle i^n x^n \rangle_x = i^n \mu_n$$

$$\Rightarrow \boxed{\mu_n = \frac{1}{i^n} \left.\frac{d^n}{dk^n} \mathcal{G}(k)\right|_{k=0}}$$

③ if moments exist : $\mathcal{G}(k) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \mu_n$

Check you calc of $\mathcal{G}(k) \to \mathcal{G}(0) = 1$ always

# Cumulant generating func.

$$\mathcal{K}(k) = \ln G(k)$$
$$= \ln \langle e^{ikx} \rangle_x$$

as with $G(k)$, we can write $\mathcal{K}(k)$ as series

$$= \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \kappa_m$$ ← replaced $\mu_m$

Cumulant $\kappa_m$ by <u>definition</u>

$$\kappa_m = \frac{1}{i^m} \frac{d^m}{dk^m} \mathcal{K}(k) \Big|_{k=0}$$

$$= \frac{1}{i^m} \frac{d^m}{dk^m} \Big[ \ln G(k) \Big] \Big|_{k=0}$$

$$\begin{tikzcd}
\text{pdf } p(x) \arrow[r, "F.T."] \arrow[dd, "\langle x^n \rangle"'] & \text{Characteristic } G(k) \arrow[d, "\ln"] \arrow[ddl, "\partial^n \ldots|_{k=0}"'] \\
 & \text{Cumulant gen. } \mathcal{K}(k) \arrow[d, "\partial^n \ldots|_{k=0}"] \\
\text{moment } \mu_n \arrow[r, dashed, "?", "\text{Bell polynom.}"'] & \text{cumulant } \kappa_n
\end{tikzcd}$$

## Relation between $\mu_m$ & $K_m$

$$G(k) = e^{\underbrace{\ln G(k)}_{\mathcal{K}(k)}}$$

replace with series

$$\sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \mu_m = \underset{\text{expansion}}{\exp}\left[ \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m \right]$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left[ \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m \right]^n$$

$k^2$ $\quad k^2 \Rightarrow K_2 \quad k \cdot k = k^2 \Rightarrow K_1^2$

$$1 + ik\mu_1 + \frac{(ik)^2}{2!}\mu_2 + \frac{(ik)^3}{3!}\mu_3 + \dots = 1 + \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m + \left( \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m \right)\left( \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} K_m \right) + \sum \cdot \sum \cdot \sum + \dots$$

collect the same powers of $k$ (with resp to factors!)

$k$ : $\mu_1 = K_1$

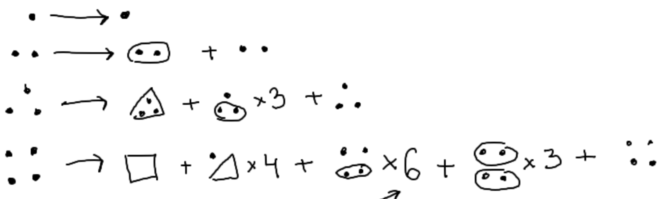
$k^2$ : $\mu_2 = K_2 + K_1^2$

$k^3$ : $\mu_3 = K_3 + K_1 K_2 \cdot 3 + K_1^3$

and so on...

$\Rightarrow$
$K_1 = \mu_1$ mean
$K_2 = \mu_2 - \mu_1^2$ variance!
$K_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3$

Clustering approach — all ways to split N dots into clusters




binomial $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

$$\mu_4 = K_4 + 4K_3K_1 + 6K_2K_1^2 + 3K_2^2 + K_1^4$$

# Correlation & Independence

A and B are indep. $\Leftrightarrow$ $P(A \cap B) = P(A) \cdot P(B)$

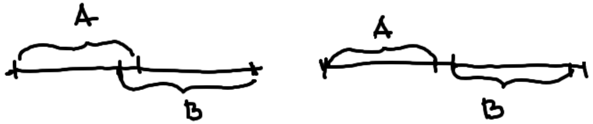
for contin. case $p(x,y) = p_x(x) \cdot p_y(y)$

$\uparrow$ marginal, $p_x(x) = \int_\Omega p(x,y')\,dy'$

multi dim case: $x_1 \dots x_n = X$

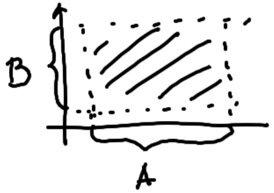
$$\mathrm{cov}(X) = \begin{pmatrix} \mathrm{cov}(x_1,x_1) & \dots & \mathrm{cov}(x_n,x_1) \\ \vdots & & \vdots \\ \mathrm{cov}(x_1,x_n) & \dots & \mathrm{cov}(x_n,x_n) \end{pmatrix} = \begin{pmatrix} \sigma^2(x_1) & \dots & \mathrm{cov}(x_n,x_1) \\ \vdots & & \vdots \\ \mathrm{cov}(x_1,x_n) & \dots & \sigma^2(x_n) \end{pmatrix}$$

1D




$\xi \in A$ and $\xi \in B$ ✗

2D




✓

covariance $\mathrm{cov}(x,y) = \mathbb{E}[(x-\langle x\rangle)(y-\langle y\rangle)]$

$$= \langle xy\rangle - \langle x\rangle\langle y\rangle$$

independence $\Rightarrow$ $\mathrm{cov}(x,y) = 0$

$\not\Leftarrow$

- Pearson corr. coef $r(x,y) = \dfrac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = \dfrac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sqrt{\langle x^2\rangle - \langle x\rangle^2}\sqrt{\langle y^2\rangle - \langle y\rangle^2}}$

- Kendall rank coef $\tau = \dfrac{[\text{number of concordant pairs}] - [\text{number of discordant pairs}]}{[\text{total number of pairs}]}$

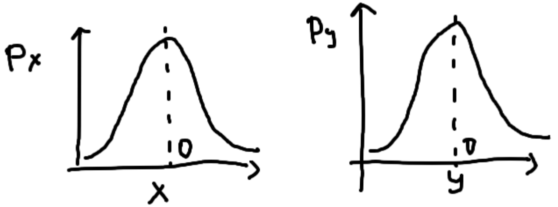
concordant — simultaneous increase or decrease in X & Y

x : 1 , 0 , 1 , 100

y : 1 , 2 , 5 , 7

discordant

**Example**: two gaussian marginal distributions and $cov(x,y)=0$




$+ \quad cov(x,y)=0$

$x \sim N(0,G)$

$W \sim \begin{cases} +1 & \text{with } p = 1/2 \\ -1 & \text{with } p = 1/2 \end{cases}$

then $Y = W \cdot x$ , W and X are indep.

Y and X are not


$$\begin{aligned} cov(x,y) &= cov(x, xW) \\ &= \langle x \cdot x \cdot w \rangle - \underbrace{\langle \cancel{x} \rangle}_{0} \langle xw \rangle \\ &= \langle x^2 \rangle \underbrace{\langle \cancel{w} \rangle}_{0} \\ &= 0 \end{aligned}$$

but $Pr\{ y > 1 \,|\, |x| < 1/2 \} = 0$

## How to check independence?

$$P(A \cap B) = P(A) \cdot P(B)$$

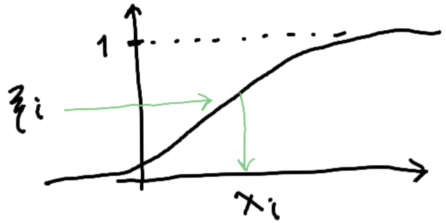
OR

$$p(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

# Sampling (HW1)

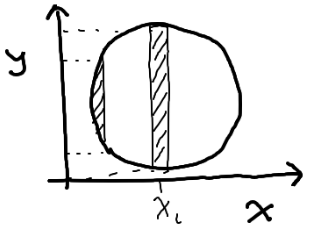
## ① Not correlated $X$ & $Y$ (OR single $X$)

1. marginalise $p_x = \int_\Omega p(x,y)\,dy$
2. get $CDF_x(x) = \int_{-\infty}^{x} p_x(x')\,dx'$
3. sample uniform $\xi_x \in Uni[0,1]$
4. generate $x_i = CDF_x^{-1}(\xi_x)$ - inverse of CDF




do the same for $y$

## ② Correlated Sampling, $X$ & $Y$

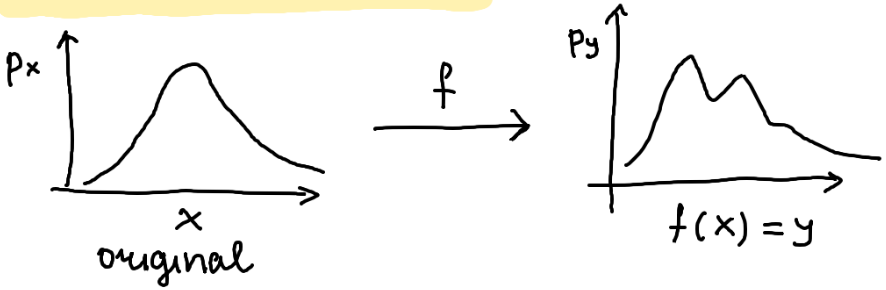



now the range of accessible values for $y_i$ depends on $x_i$

1. first steps
2. are the same
3.
4. generate $x_i$
5. compute $p(y|x) = \dfrac{p(x,y)}{p_x(x)}$
6. get conditional CDF $F(y|x) = \int_{-\infty}^{y} p(y'|x)\,dy'$
7. sample $\xi_y \in Uni[0,1]$
8. for given $x_i$ generate $y = F^{-1}(\xi_y|x_i)$

direct correlated sampling may be tedious due to inverse functions

# PDF transformation




exact mapping of CDFs should be guaranteed $\quad P(X = x) = P[f(X) = f(x)] = P(Y = y)$

$$F_y(y) = F_x(x)$$

take derivatives: $\frac{d}{dx} F_y(y) = F_y'(y) \frac{dy}{dx} = p_y(y) \frac{dy}{dx} = p_x(x)$

since pdf $\geq 0$ we take abs. val

$$p_y(y) = \frac{1}{\left|\frac{dy}{dx}\right|} p_x(x)$$

$$= \left|\frac{df(x)}{dx}\right|^{-1} p_x(x)$$

multivariate $\quad x = (x_1 \ldots\ldots x_n)$

$f: \mathbb{R}^n \to \mathbb{R}^n$

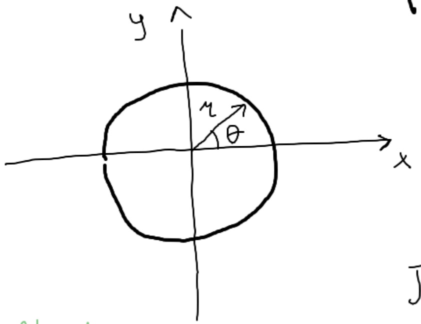
$$J_T = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

Jacobian matrix

shows how much space is expanded/contracted by a transform.

$$p_y(y) = \frac{1}{|\det J_T|} p_x(x)$$

# Example: polar transform. sampling from uniform disk




Check:

$r = \xi_1$

$\theta = 2\pi \xi_2$

$\xi_1\, \xi_2 \in \mathrm{Uni}[0,1]$

(1) $x = r \cdot \cos\theta$
$y = r \cdot \sin\theta$

$$\begin{pmatrix} x \\ y \end{pmatrix} \xleftarrow{f} \begin{pmatrix} r \\ \theta \end{pmatrix}$$

$$J_T = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$\det J_T = r(\cos^2\theta + \sin^2\theta) = r$ , $r > 0$ no need to take abs

we want to have uniform density in $(x,y)$ so $p(x,y) = \frac{1}{\pi R_d^2} = \frac{1}{\pi}$ , $R_d = 1$

then $p(x,y) = \frac{p(r,\theta)}{r} = \frac{1}{\pi}$

$\Rightarrow p(r,\theta) = \frac{r}{\pi}$

1. $p_r(r) = \int_0^{2\pi} p(r,\theta)\, d\theta = 2r$

2. $\mathrm{CDF}_r(R) = \int_0^R 2r\, dr = R^2$

3, 4 $\Rightarrow$ produce $r_i$

5. $p(\theta \mid r) = \frac{p(r,\theta)}{p(r)} = \frac{r}{\pi} \frac{1}{2r} = \frac{1}{2\pi}$ — $p(r,\theta) = p(r)p(\theta)$ independent

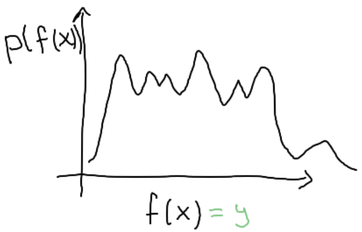
6. $F(\theta \mid r) = \int_0^\theta \frac{d\theta}{2\pi} = \frac{\theta}{2\pi}$ ← uniform

7, 8 $\Rightarrow$ produce $\theta$ from uni$[0, 2\pi]$

now we can sample $(x,y)$ coord. using $(r,\theta)$ pairs.

# Function composition & Normalising flows


$$\underbrace{f_n(\overbrace{f_{n-1} \cdots f_1(x)}^{y_1})}_{y_{n-1}} \longrightarrow$$




$f = f_n \circ \cdots f_i \cdots \circ f_1$ - composition:

1. $f_i$ - easily invertible
2. $J_i$ - easy to compute

$$p_x(x) = \left| \det \frac{d f(x)}{d x} \right| \cdot p_y(y)$$

$$= \left| \det \frac{d f_n}{d y_{n-1}} \cdot \frac{d f_{n-1}}{d y_{n-2}} \cdots \frac{d f_1(x)}{d x} \right| \cdot p_y(y)$$

$$= \left| \det \left( \prod_{i=1}^{n} \frac{d f_i}{d y_{i-1}} \right) \right| p_y(y)$$

$$= \prod_{i=1}^{n} \left| \det \frac{d f_i}{d y_{i-1}} \right| p_y(y)$$

lets apply log: $\prod \to \sum$

$$\log p_x(x) = \sum_{i=1}^{n} \log \left( \left| \det \frac{d f_i}{d y_{i-1}} \right| \right) + \log p_y(f(x))$$

$$\text{loss} = -\frac{1}{|\text{Dataset}|} \sum_{y \in D} \log p_y(f(x))$$

negative log-likelihood (NLL)

finally

$$\underbrace{\log p_y(f(x))}_{\text{output / loss}} = \underbrace{\log p_x(x)}_{\text{input}} - \underbrace{\sum_{i=1}^{n} \log \left( \left| \det \frac{d f_i}{d y_{i-1}} \right| \right)}_{\text{flow}}$$

1 Inference

x

Data space

$y = f(x)$ ⟶

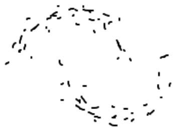
y Gaussian noise

latent space

example taken from „Real NVP" paper

2 Generation

flow-based generative model



⟵ $x = f^{-1}(y)$

# Copulas

„link" in latin

$Copula(CDF_1, CDF_2 \dots, \alpha) \longrightarrow$ Joint distribution

takes on input marginal CDFs and
some parameters, produces joint distr.

we want somehow add/mix $F_1$ & $F_2$:

$P(A) + P(B) > 1$ - prohibited

$F_1, F_2$ are in $[0,1]$ domain

$\Psi$ transform them into different domain

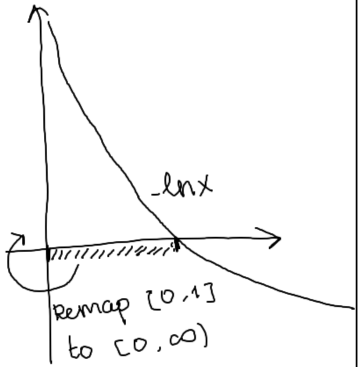
$\Psi(F_1), \Psi(F_2) \subset [0, \infty)$

perform some operations

$\Psi^{-1}$ return them back

$\Psi^{-1}(\Psi(F_1) + \Psi(F_1)) \subset [0, 1]$

$\Psi$ - generator func




$\Psi = -\ln x$

$\Psi^{-1} = e^{-x}$

$C(P_A, P_B) = e^{-[(-\ln P_A) + (-\ln P_B)]} = P_A \cdot P_B$

(quite useless independent copula :))

But if we introduce param $\alpha$ like

$\Psi[F(x)] = [-\ln F(x)]^{\alpha}$

we'll get Gumbel Copula:

$Gumbel(F_1, F_2, \alpha) = e^{-[(-\ln F_1)^{\alpha} + (-\ln F_2)^{\alpha}]^{1/\alpha}}$

where $\alpha = 1 \sim$ independence

$\alpha \to \infty \sim$ fully correlated

(50 is enough)