Introduction to Statistics II

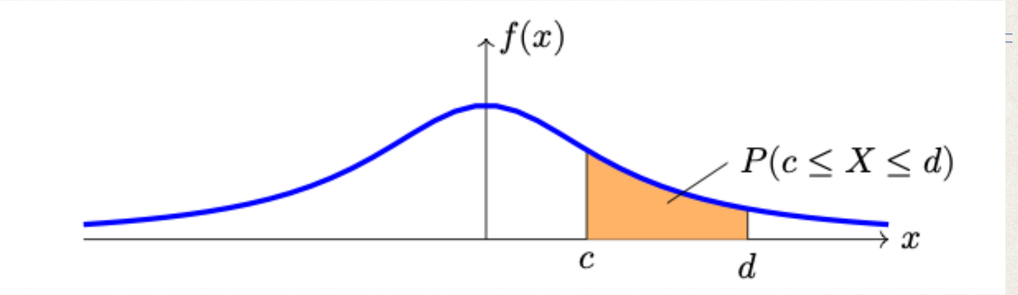
Atreyee Sinha

Review:

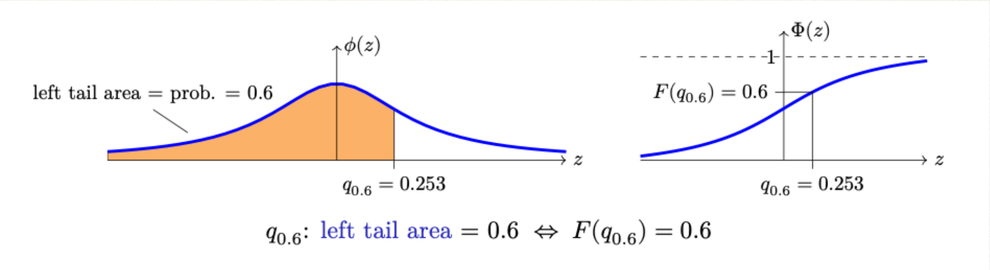
- Last lecture:
  - Basic probability, Bayes theorem, Discrete probability distributions
- This lecture:
  - Continuous distributions
  - Variable transformations
  - Random number generations
  - Central limit theorem
  - Joint pdf and marginal pdf
  - Correlations

Continuous random variables

- $P(c \leq X \leq d) = \int_c^d f(x)dx$



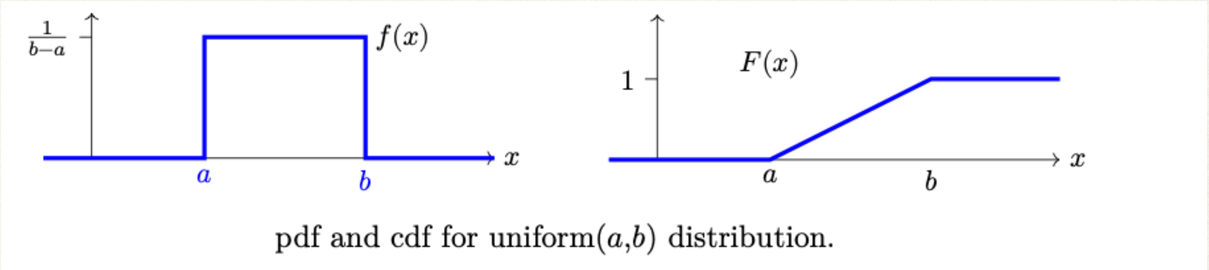
- note: $P(X = a) = 0$; always defined between two intervals
- CDF: $F(b) = P(X \leq b) = \int_{-\infty}^b f(x)dx$
- $E[X] = \int xf(x)dx$



- Properties of mean and variance same as that of the discrete case
- Quantiles/percentiles:
  - The pth quantile of X is the value q_p such that $P(X \leq q_p) = p$.

Uniform distribution

- $f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$
- $F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases}$



pdf and cdf for uniform(a,b) distribution.

- $E[X] = \frac{a+b}{2}$; $\text{Var}(X) = \frac{(b-a)^2}{12}$
- Eg: Angle of a spinning wheel,

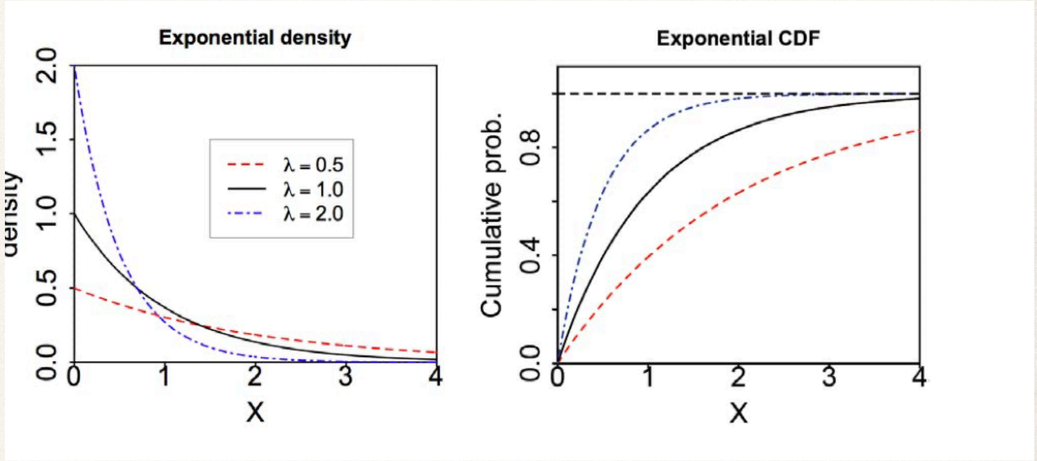
Exponential Distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

$$E[X] = \frac{1}{\lambda}$$

$$\text{Var}(X) = \frac{1}{\lambda^2}$$



- Eg: Independent events at constant average rate
  - # particles detected per hour
  - # photons in a given pixel
  - # waiting time for a Uber (not a metro!)

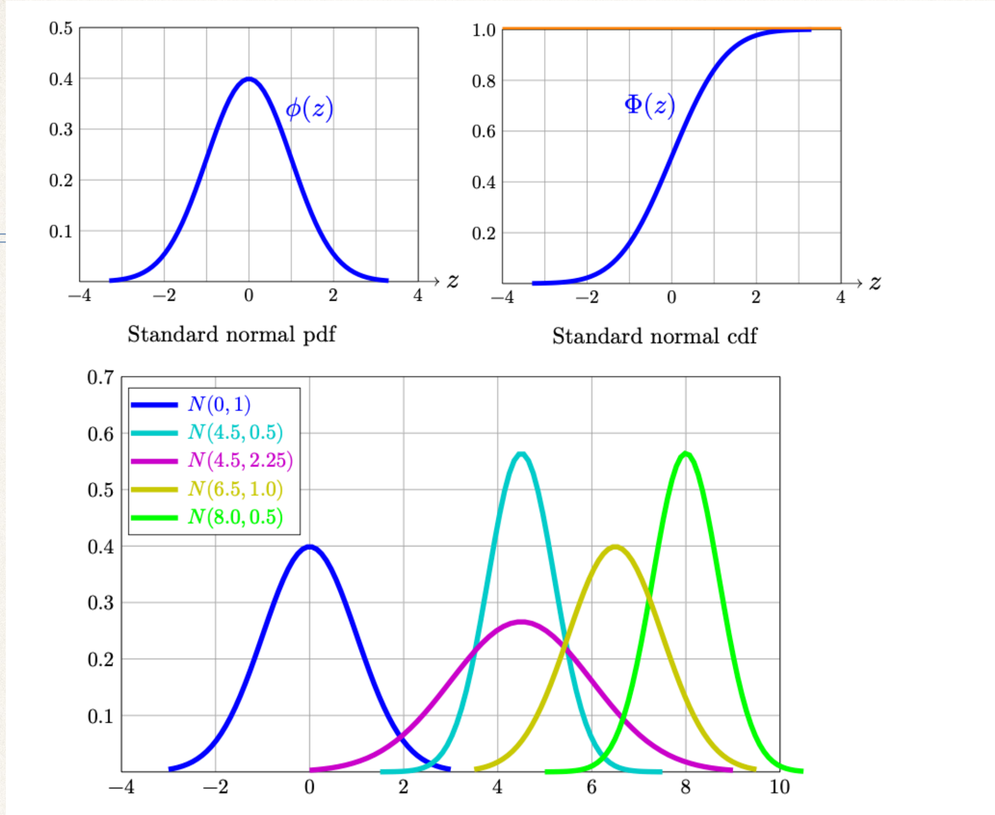
Gaussian distribution

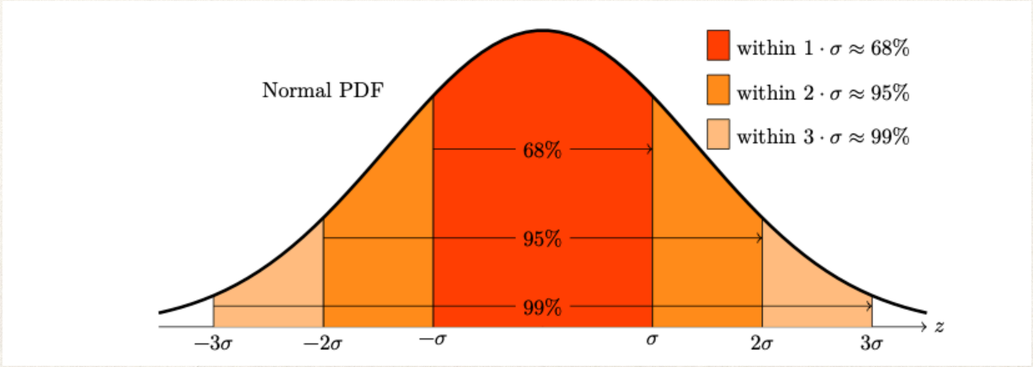
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$E[X] = \mu$$

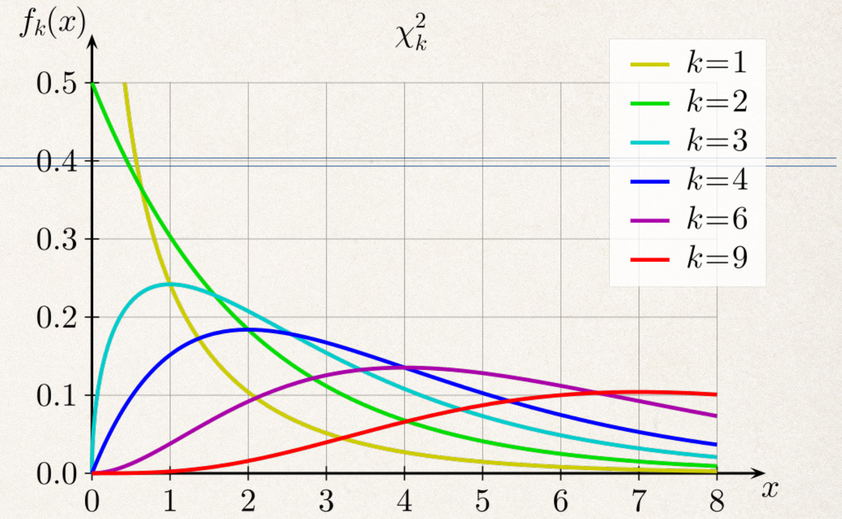
$$\operatorname{Var}(X) = \sigma^2$$

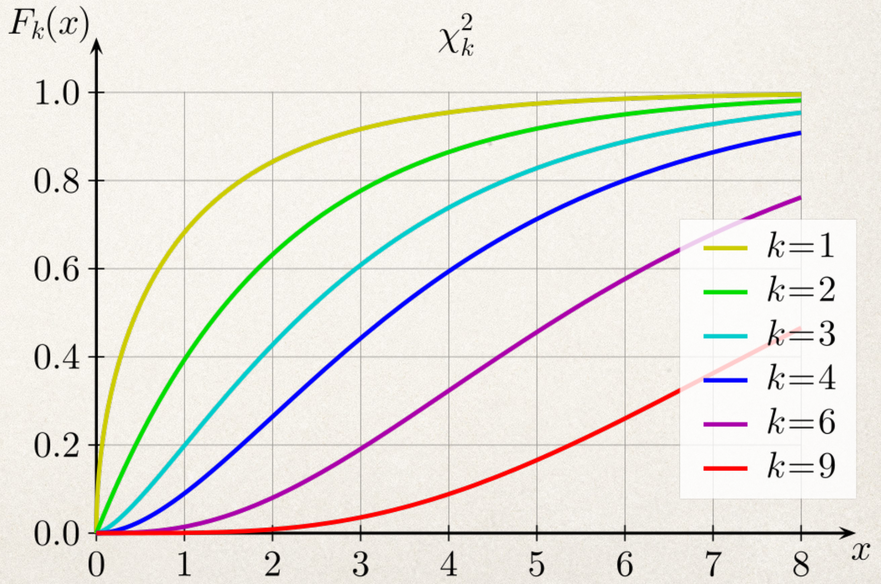




Chi-sq distribution

- A continuous probability distribution defined strictly by degrees of freedom, denoted as k
- It represents the sum of squared independent standard normal random variables.
- Widely used in hypothesis testing
- $$f(x;k) = \begin{cases} \dfrac{x^{(k/2)-1}e^{-x/2}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$
- $$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$$





Transformations of random variables

- Goal:
  - Given PDF $f_X(x)$ and a transformation function $y = g(x)$, find the PDF of the new variable $f_Y(y)$
- Method:
  - Find the CDF:
  - $F_Y(y) = P(g(X) \leq y)$.
  - Differentiate the CDF with respect to y to get the PDF: $f_Y(y) = \frac{dF_Y(y)}{dy}$

- Example $Y = X^2$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right]$$

If $X \sim N(0,1)$; $f_Y(y) \sim = \frac{1}{\sqrt{y}} f_X(\sqrt{y})$.

A word about random number generations

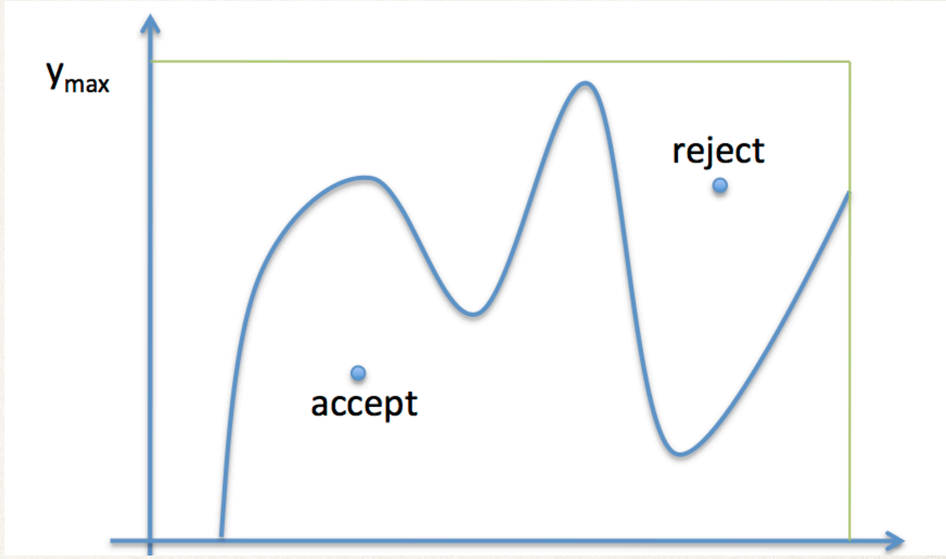
- How are random numbers generated?
  - Pseudo Random numbers (PRNG) :
    - Initial seed with sets of mathematical functions
      - Eg: linear shift bit registers, $X_{n+1} = (a \cdot X_n + c) \mod m$, etc
    - Same seed -> same number (repeatable)
    - High level languages can generate seeds (eg: system time, some clock state)
    - Usually Uniformly distributed
  - True Random numbers (TRNG):
    - Rather data from “true” unpredictable systems; thermal fluctuations, electronic noise…

Generating distributions of any pdf

- Lets say you want to generate random number for any arbitrary pdf. How will you approach it?
- Transform a Uniform Random Distribution into any given distribution
  - Find the CDF: Determine the target continuous distribution's Cumulative Distribution Function $F(x) = P(X \leq x)$.
  - Invert the CDF: Solve the equation $u = F(x))\ for\ (x)$ to isolate the inverse function $(x = F^{-1}(u))$.
  - Plug in Uniform Values: Generate a uniform random number u from $Unif(0,1)$ and plug it into $F^{-1}(u)$.
  - Works well if the target CDF has closed form analytic expression. If not, use additional transformations

Rejection Sampling

- The Problem
  - Target PDF $f(x)$ is too complex to invert directly.
- The Solution
  - Use a simpler, known proposal distribution $g(x)$.
- Scale $g(x)$ by a constant M to fully envelope $f(x)$.
- Constraint:
  - $f(x) \leq M \cdot g(x)$ *for all* x.
- Efficiency Metric
  - Efficient choice of initial function is useful



The central limit theorem

The Statement

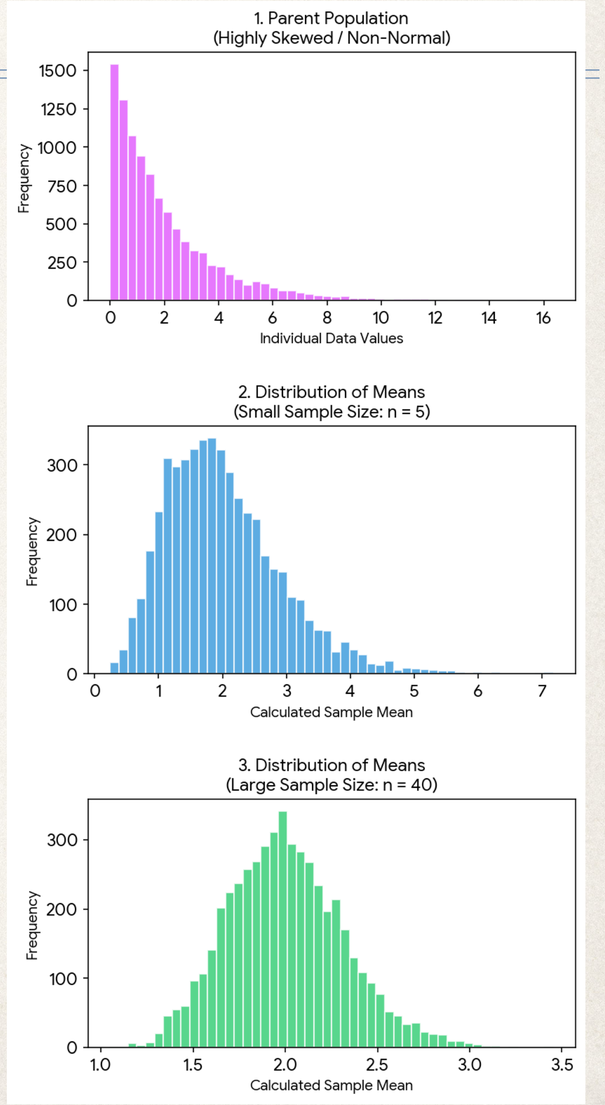
- The Central Limit Theorem (CLT) states that the **distribution of the sample mean (or sum)** approaches a **normal distribution** as the sample size n becomes larger, *regardless of the shape of the underlying population distribution*.
- Restated: If a population has a mean μ and a finite variance σ^2, then the standardized variable Z converges to a standard normal distribution:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

- Assumptions:
  - Independent and identically distributed variables
  - Underlying distribution must have a well defined mean and variance
  - Valid for large sample size (how large?)

Mechanism

- Think of a highly skewed, distribution
- Individual data points from this population do not look like a normal curve.
- Think of a highly skewed, flat, or dual-peaked distribution (like a rolling standard 6-sided die).
- Individual data points from this population do not look like a normal curve.
- Plotting all collected sample means always results in a symmetric, bell-shaped distribution.
- Center: The average of all sample means converges exactly to the population mean μ.
- Spread: The standard deviation of these means shrinks as sample size grows, calculated as $\sigma/\sqrt{n}$.

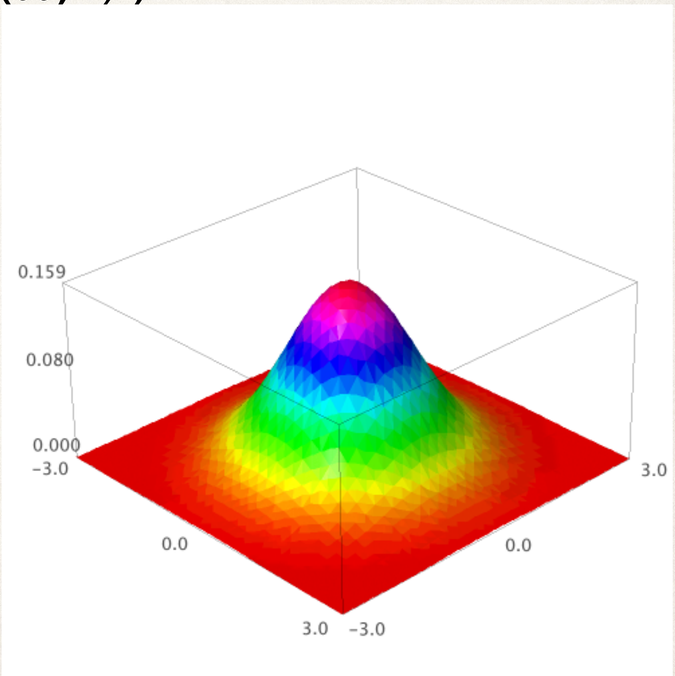


Extremely powerful

- Most real wold distributions are complex
- The CLT acts as a statistical bridge, allowing researchers to study populations without knowing their exact native distributions.
- Explains why we “mostly” end up with Gaussian data

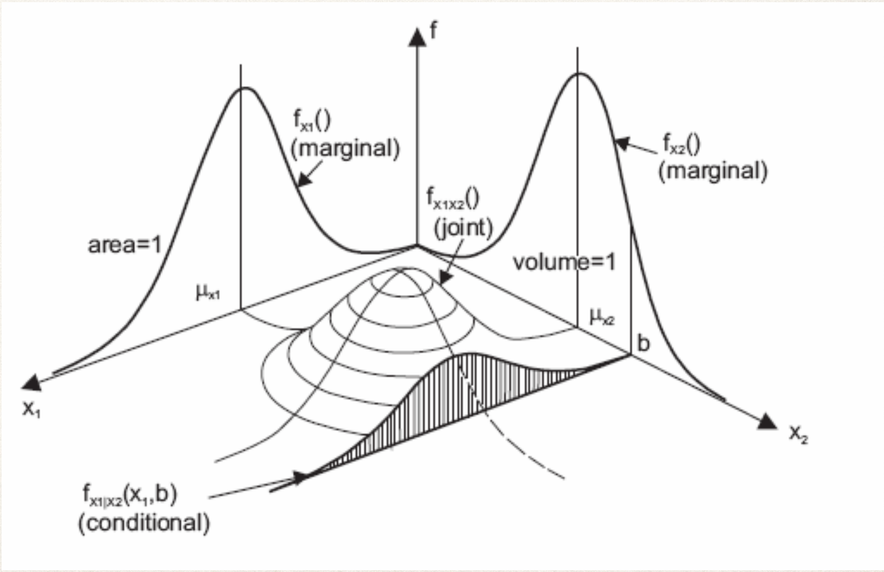
Joint distributions

- The joint probability density function (joint pdf) of X and Y is a function $f(x, y)$ giving the probability density at (x, y)
  - $0 \leq f(x, y)$
  - The total probability is 1
- $P(a \leq X \leq b; c \leq Y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$



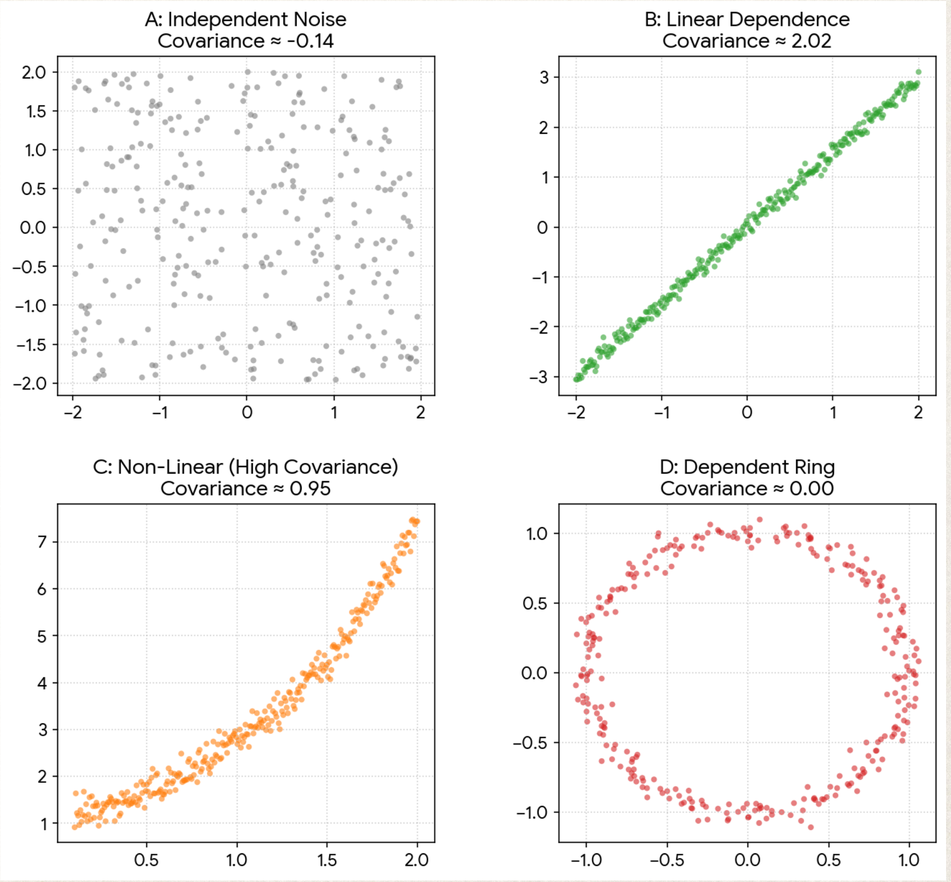
Marginal distributions

- Marginal pdf: Collapses one variable to study the other alone.
  - $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$
- Independence: $f(x, y) = f_X(x) \cdot f_Y(y)$



Covariance and correlations

- How much do two random variable vary together?
  - Population cov: $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
  - If X and Y are independent then $\text{Cov}(X, Y) = 0$.
    - The converse is false: zero covariance does not always imply independence
- Correlations: "remove scale from the covariance"
  - $\text{Cor}(X, Y) = \dfrac{Cov(X, Y)}{\sigma_X \sigma_Y}$
  - Dimensionless



Accuracy vs Precision

- Large sample sizes —> better precision (=/=> better accuracy)
- Accuracy:
  - Closeness to the true value
  - Systematic errors
- Precision:
  - ~ Variance, reproducibility
  - Random errors
- Eg:
    - OPERA neutrino anomaly (High precision, low accuracy)
    - Eddington's test of GR (Accurate but imprecise)