Physics 509 Theory of Measurements Course Notes

Tobias Faehndrich

This document was typeset on September 27, 2025

Introduction:
Notes written at UBC 2025W1 with Dr. Colin Gay.

Contents

1 Foundations of Probability Theory

This lecture covered the course structure and grading.

1.1 Motivation: Stochastic Nature of Experimental Data

- Stochastic processes:
  - muon decay
  - inherent stochasticity
  - quantum mechanics
- Mostly concerned with measurement devices — how do we measure?
- Example: a muon lifetime experiment
  - Take a cosmic muon, detect light, and discriminate.
  - Muon decays into an electron and neutrinos, and the electron produces light.
  - Measure the time between light pulses.
  - Many factors cause noise in the data — results change even if the same mechanism occurs twice.

1.2 Probabilistic Interpretation of Experimental Results

- Experiments are repeated trials.
- Probability (probabilistic interpretation):
  - Results are interpreted as the long-term average of repeating an experiment many times.
  - Example: coin flip

$$P(H) = \lim_{N \to \infty} \frac{n_H}{N}$$

$$n(H) = \text{number of heads in } N \text{ trials}$$

1.3 Sample Spaces and Stochastic Variables

- In modern probability theory:
  - 3 axioms (Kolmogorov)
  - Let X be a stochastic variable.
  - Define sample space S (Ω):

$$S = \{x_1, x_2, ...\}$$

  - Examples:
    1. Coin flip:
    $$S = \{H, T\}$$
    2. Roll a die:
    $$S = \{1, 2, 3, 4, 5, 6\}$$
    3. Grade in this class:
    $$S = \{0, 1, 2, ..., 100\}$$
    4. Decay time of a radioactive atom:
    $$S = [0, \infty)$$
  - S can be finite (Binomial), countable (Poisson), or infinite (Gaussian, Uniform).

1.4 Events and Set Operations

- Definition: An event E is a subset of S.
- Example: one die roll

$$S = \{1, 2, 3, 4, 5, 6\}$$
$$E = \text{rolling an even number} = \{2, 4, 6\}$$

- Example: E = atom decayed by time t_0

$$S = [0, t_0]$$

- Operations on events:
  - Union (OR) and Intersection (AND)
  - Let A, B be events in S:

$$E = A \cup B = \{e : e \in A \text{ or } e \in B \text{ (or both)}\}$$

  - Example: flip a coin twice

$$S = \{HH, HT, TH, TT\}$$
$$A = \text{1st flip is H} = \{HH, HT\}$$
$$B = \text{2nd flip is H} = \{HH, TH\}$$
$$A \cup B = \{HH, HT, TH\}$$
$$A \cap B = \{e \mid e \in A \text{ and } e \in B\} = \{HH\}$$
$$AB = A \cap B$$
$$A^c = \{e \mid e \in S \text{ and } e \notin A\} = \{TH, TT\}$$

- Properties:
  - Commutative:

$$A \cup B = B \cup A, \quad AB = BA$$

  - Associative:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad (AB)C = A(BC)$$

  - Distributive:

$$(A \cup B)C = AC \cup BC, \quad A(B \cup C) = AB \cup AC$$

  - De Morgan's Laws:

$$(A \cup B)^c = A^c B^c, \quad (AB)^c = A^c \cup B^c$$

1.5 Kolmogorov's Axioms of Probability

- A function P on S is a probability measure if it satisfies:
  1. $P(S) = 1$
  2. $P(\emptyset) = 0$
  3. For any countable sequence of disjoint events $E_1, E_2, ...$ in S:

$$E_i E_j = \emptyset \text{ for } i \neq j$$
$$P\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

1.6 Consequences of the Probability Axioms

- $$P(\emptyset) = 0$$

  Let

  $$E_1 = S, \quad E_2 = \emptyset$$
  $$E_1E_2 = \emptyset$$
  $$P(S \cup \emptyset) = P(S) + P(\emptyset) = 1 + P(\emptyset)$$
  $$P(S) = 1,\, P(\emptyset) = 0$$

- $$P(E^c) = 1 - P(E)$$
  $$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

- If $B \subset A$, then:

  $$P(B) \leq P(A)$$
  $$A = B \cup (B^cA)$$
  $$P(A) = P(B \cup (B^cA))$$
  $$P(B) = P(A) - P(B^cA) \leq P(A)$$

- $$P(A \cup B) = P(A) + P(B) - P(AB)$$

  If we let the areas of the Venn diagram be 1 (A), 2 (A+B), 3 (B), then:

  $$A \cup B = 1 \cup 2 \cup 3$$
  $$P(A \cup B) = P(1 \cup 2 \cup 3) = P(1) + P(2) + P(3)$$
  $$P(A) = P(1) + P(2), \quad P(B) = P(2) + P(3)$$
  $$P(A) + P(B) - P(2) = P(1) + P(2) + P(3) = P(A \cup B)$$
  $$\text{equivalently } P(A) + P(B) - P(AB) = P(A) + P(B) - P(AB)$$

1.7 Uniform Probability on Finite Sample Spaces

- $$E_i = S_i \text{ for } i = 1, 2, ...n$$
  $$E_iE_j = \emptyset \text{ for } i \neq j$$
  $$S = \cup_{i=1}^n E_i$$
  $$P(S) = 1 = P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$$
  $$P(E_i) = P(E_j) \quad \text{all equally likely}$$
  $$1 = \sum_{i=1}^N P(E_i) = NP(E_i)$$
  $$P(E_i) = \frac{1}{N} = P(E_j)$$
  $$N = |S| = \text{number of elements in (cardinality of) } S$$
  $$F \text{ be any event (set) in } S \text{ with } k \text{ elements} \quad |F| = k$$
  $$P(F) = P(\cup_{S_i \in F}\{E_i\}) = \sum_{i=1}^k P(E_j) = \sum_{i=1}^k \frac{1}{N} = \frac{k}{N} = \frac{|F|}{|S|}$$

1.8 Example: Probability of a Straight in Poker

- Example: 5-card poker hand forming a straight

$$S = \{(AC, 2C, 3C, 4C, 5C), (2C, 3C, 4C, 5C, 6C), ...\}$$

$$S = \binom{52}{5} = \frac{52!}{5!\,47!} = 2,598,960$$

- Event = straight = 5 consecutive cards, not of the same suit, any starting card.

$$10(4^5 - 4) = 10200$$

- Starting cards: Ace (A,2,3,4,5), 2 (2,3,4,5,6), ..., 10 (10,J,Q,K,A)
- Not all the same suit: $4^5 - 4$ (exclude all same suit)

$$P(\text{straight}) = \frac{10(4^5 - 4)}{\binom{52}{5}} = 0.00392465$$

1.9 Conditional Probability

- Given 2 events E, F, sample space S:

$$P(E) = \text{probability of a trial from } S \text{ in } E$$

$$P(F) = \text{probability of a trial from } S \text{ in } F$$

- Conditional probability of E given F has occurred:

$$P(E|F) = \text{probability of a trial from } S \text{ in } E, \text{ given the trial is in } F$$

- Note: $P(EF)$ is the probability of a trial from S in both E and F.
- Need to normalize by $P(F)$, so we define:

$$P(E|F) = \frac{P(EF)}{P(F)} \quad \text{if } P(F) > 0$$

$$P(EF) = P(E|F)P(F)$$

- Example: flip a coin 2 times

$$S = \{HH, HT, TH, TT\}$$

  Conditional probability of $HH \equiv A$ given:

  - First flip $= H \equiv B = \{HH, HT\}$
  - Either flip is $H \equiv C = \{HH, HT, TH\}$

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(\{HH\})}{P(\{HH, HT\})} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(A|C) = \frac{P(AC)}{P(C)} = \frac{P(\{HH\})}{P(\{HH, HT, TH\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

2 Bayesian Inference and Random Variables

2.1 Bayes' Formula

- Let E, F be events:

$$E = EF \cup EF^c$$
$$P(E) = P(EF) + P(EF^c)$$
$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$
$$P(E) = P(E|F)P(F) + P(E|F^c)(1 - P(F))$$

- **Example:** Suppose a blood test is 95% effective in detecting a disease if the person has it. It also has a 1% false positive rate. Suppose 0.5% of the population has the disease.

$$D = \text{person has disease}$$
$$E = \text{test is positive}$$

- We want:

$$P(D|E) = \frac{P(ED)}{P(E)}$$
$$P(D|E) = \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)(1 - P(D))}$$
$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} = 0.32$$

- So even with a positive test, there is only a 32% chance of having the disease.

2.2 Law of Total Probability

- Let $\{F_i\}$ be mutually exclusive events such that:

$$\cup_{i=1}^{n} F_i = S$$

  Then for any event E:

$$E = E \cap \left(\cup_{i=1}^{n} F_i\right) = \cup_{i=1}^{n} (EF_i)$$
$$P(E) = P\left(\cup EF_i\right) = \sum_{i=1}^{n} P(EF_i) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$$

2.3 Independent Events

- Generally, $P(E|F) \neq P(E)$.
- If knowing F does not change the probability of E:

$$P(E|F) = \frac{P(EF)}{P(F)} = P(E)$$
$$\boxed{P(EF) = P(E)P(F)}$$

2.3.1 Example: Rolling Two Dice

- Let:

$$E_1 \equiv \text{sum} = 6$$
$$F \equiv \text{first die} = 4$$
$$E_1 : \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$
$$F : \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$$
$$E_1F = \{(4,2)\}$$
$$P(E_1F) = \frac{1}{36}$$
$$P(E_1) = \frac{5}{36}$$
$$P(F) = \frac{6}{36} = \frac{1}{6}$$
$$P(E_1)P(F) = \frac{5}{36} \times \frac{1}{6} = \frac{5}{216} \neq P(E_1F)$$

- Let:

$$E_2 \equiv \text{sum} = 7$$
$$E_2 : \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$
$$E_2F = \{(4,3)\}$$
$$P(E_2) = \frac{6}{36} = \frac{1}{6}$$
$$P(F) = \frac{1}{6}$$
$$P(E_2F) = \frac{1}{36}$$

2.4 Random Variables and Probability Distributions

- $S = \{\text{all possible outcomes of stochastic process } X\}$

$$x = \text{random variable}$$

$$S = \text{finite or countable infinite: discrete random variable}$$
$$S = \text{uncountable infinite: continuous random variable}$$

- Continuous case:

$$P(x_0, x_0 + dx) = p(x)dx$$

where $p(x)$ is the probability density function (pdf).

- Discrete case:

$$S = S_i$$
$$p_i = \text{probability of } S_i \quad \text{(probability mass function, pmf)}$$

$$0 \leq P(S_i) \leq 1$$
$$1 = P(S)$$
$$0 \leq p(x)$$
$$\int_{-\infty}^{\infty} p(x)dx = 1$$

2.5 Describing a Distribution

- To describe $p(x)$ in general we specify:
  - **Mode —** peak value of $p(x)$
  - **Median —** 50% cumulative value
  - **Mean —** average value of x weighted by $p(x)$

2.6 Cumulative Distribution Function (CDF)

- $$F(x) = \int_{-\infty}^{x} p(x')dx' = P(X \leq x)$$
  $$F(-\infty) = 0, \quad F(\infty) = 1$$

2.7 Expectation Values

- Expectation of any function $f(x)$ over $p(x)$:

$$E(f) = \int_{\Omega} f(x)p(x)dx$$

E is a linear operator: $E(af + bg) = aE(f) + bE(g)$

- Expectation of powers of x:

$$E(x^0) = E(1) = \int 1 \cdot p(x)dx = 1$$

$$E(x^1) = \int xp(x)dx \equiv \mu = \text{mean value of } x$$

$$E(x^2) = \int x^2 p(x)dx \equiv \sigma^2 = \text{variance of } x$$

2.8 Characteristic Function

- The characteristic function of $p(x)$:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} p(x)dx = E(e^{itx})$$

$$\varphi(t) = E\left(1 + itx + \frac{(itx)^2}{2!} + \ldots\right)$$

$$= 1 + itE(x) + \frac{(it)^2}{2!}E(x^2) + \ldots$$

$$= \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mu_{k'}$$

- Moments from $\varphi(t)$:

$$\left.\frac{d^n \varphi(t)}{dt^n}\right|_{t=0} = i^n \mu_{n'}$$

2.9 Central Moments

- $$E((x-\mu)^n) = \int (x-\mu)^n p(x) dx \equiv \mu_n$$
  $$\mu = E(x)$$
- 1st central moment:
  $$E((x-\mu)^1) = E(x) - E(\mu) = \mu - \mu = 0$$
- 2nd central moment (variance):
  $$E((x-\mu)^2) \equiv V(x) = \sigma^2$$
- 3rd central moment (skewness):
  $$\text{skewness} = \frac{E((x-\mu)^3)}{\sigma^3}$$
- 4th central moment (kurtosis):
  $$\text{kurtosis} = \frac{E((x-\mu)^4)}{\sigma^4} - 3$$
  (The -3 ensures that the kurtosis of a normal distribution is 0.)

3 Bayesian Reasoning and Probability Distributions

Tuesday, September 16th 2025

3.1 Bayes Theorem and Its Applications

- Bayes Theorem: for events A and B, we have

$$\boxed{P(AB) = P(A|B)P(B) = P(B|A)P(A) = P(BA)}$$

- Usually it is given in this form:

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

- People argued about when you are allowed to use this theorem.

3.2 The Monty Hall Problem: A Bayesian Analysis

- Example: Monty Hall Problem (Game show with host named Monty Hall)
  - There are 3 doors; behind one is a car, behind the other two are goats.
  - You select a door; if the car is behind it, you win.
  - Twist: after you select a door, Monty opens one of the other 2 doors to reveal a goat.
  - Question: stay or switch?
  - Solution: use Bayes theorem.
  - Sample space: $S = \{C_1 = \text{cgg}, C_2 = \text{gcg}, C_3 = \text{ggc}\}$
  - Event 2 = MH opens door 2.
  - Event 3 = MH opens door 3.
  - Number such that your choice is door 1.
  - Take case E_2, then we want to know $P(C_1|E_2)$.

$$P(C_1|E_2) = \frac{P(E_2|C_1)P(C_1)}{P(E_2)}$$

  - $P(C_1) = \frac{1}{3}$
  - $P(E_2|C_1) = \frac{1}{2}$ because if the car is behind door 1, Monty can open either door 2 or 3.
  - $P(E_2) = \frac{1}{2}$
  - Law of total probability:

$$P(E_2) = P(E_2|C_1)P(C_1) + P(E_2|C_2)P(C_2) + P(E_2|C_3)P(C_3) = \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{1}{2}$$

  - $P(C_1|E_2) = \frac{P(E_2|C_1)P(C_1)}{P(E_2)} = \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}$
  - $P(C_1|E_2) = \frac{1}{3}$
  - $P(C_2|E_2) = 0$
  - $P(C_3|E_2) = \frac{P(E_2|C_3)P(C_3)}{P(E_2)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

3.3 Alternate Monty Hall Formulations

- Alternate version: $E =$ MH shows you a goat from $\{2, 3\}$.
  - We want to find $P(C_1|E)$.
  - $P(C_1|E) = \frac{P(E|C_1)P(C_1)}{P(E)}$
  - $P(C_1) = \frac{1}{3}$
  - $P(E|C_1) = 1$ because if the car is behind door 1, Monty can open either door 2 or 3.
  - $P(E) = 1$ by law of total probability:

$$P(E) = P(E|C_1)P(C_1) + P(E|C_2)P(C_2) + P(E|C_3)P(C_3) = 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 1$$

  - $P(C_1|E) = \frac{1 \cdot \frac{1}{3}}{1} = \frac{1}{3}$
- Another version: What if MH does *not* know where the car is?
  - $E =$ MH opens $\{2, 3\}$ and reveals a goat.
  - We want to find $P(C_1|E)$.
  - $P(C_1|E) = \frac{P(E|C_1)P(C_1)}{P(E)}$
  - $P(C_1) = \frac{1}{3}$ because we picked door 1.
  - $P(E|C_1) = \frac{1}{2}$ because if the car is behind door 1, Monty can open either door 2 or 3 since he does not know where the car is.
  - By law of total probability:

$$P(E) = P(E|C_1)P(C_1) + P(E|C_2)P(C_2) + P(E|C_3)P(C_3) = 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{2}{3}$$

  - $P(C_1|E) = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$

3.4 Monty Hall Generalized to n Doors

- Now back to the standard version of the problem but with n doors.
  - You pick door 1, MH opens any door with a goat behind it from 2 to n ($n - 1$ options).
  - $P(C_1|E) = \frac{P(E|C_1)P(C_1)}{P(E)}$
  - $P(E) = 1$ because he can always choose a door with a goat behind it (many options and he knows the answers).
  - $P(C_1) = \frac{1}{n}$
  - $P(E|C_1) = 1$ because if the car is behind door 1, Monty can open any of the other doors.

3.5 Continuous Probability Distributions and Moments

- Continuous probability distribution $p(x)$:
- Moments:

$$E(x^n) = \int_{-\infty}^{\infty} x^n p(x)\, dx$$

mean:	$\mu = E(x)$
variance:	$V(x) = \sigma^2 = E((x-\mu)^2) = E(x^2) - \mu^2$
std dev:	$\sigma = \sqrt{\sigma^2}$

- Central moments:

$$E(x-\mu) = E(x) - \mu = 0$$
$$E((x-\mu)^2) = \sigma^2$$
$$E((x-\mu)^3) = \text{skewness}$$
$$E((x-\mu)^4) = \text{kurtosis}$$

- Characteristic function:

$$\Phi(t) = E(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} p(x)\, dx \tag{3.1}$$
$$= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mu_k \tag{3.2}$$

$$\Phi_\mu(t) = E(e^{it(x-\mu)}) = E(e^{itx})e^{-it\mu} = \Phi(t)e^{-it\mu}$$

$$V(x) = E((x-\mu)^2) \tag{3.3}$$
$$= E(x^2 - 2\mu x + \mu^2) \tag{3.4}$$
$$= E(x^2) - 2\mu E(x) + \mu^2 E(1) \tag{3.5}$$
$$= E(x^2) - 2\mu^2 + \mu^2 \tag{3.6}$$
$$= E(x^2) - \mu^2 = E(x^2) - (E(x))^2 \tag{3.7}$$

3.6 Discrete Probability Distributions

- The discrete case (e.g., rolling a die, picking a card) uses a probability mass function.
- Usually denote outcomes as r:
- p_r = probability of outcome r.
- $\sum_r p_r = 1$
- $E(r) = \sum_r p_r r = \text{mean } \mu$
- Variance: $V(r) = \sum_r (r-\mu)^2 p_r = E(r^2) - \mu^2$
- Coin flip example: $S = \{H, T\}$.
- Often map to 0 or 1: $H = 0$, $T = 1$.
- But in theory you can pick any two numbers a and b to map outcomes, just so you can calculate mean and variance.

$$E(r) = ap_H + bp_T$$

3.7 Cumulative Distribution Functions

- For continuous case:
$$F(x) = \int_{-\infty}^{x} f(x')\,dx'$$
- For discrete case:
$$F(r) = \sum_{r' \leq r} p_{r'}$$
- $F(x)$ is the cumulative distribution function (CDF).
- $F(x)$ is non-decreasing, $F(-\infty) = 0$, $F(\infty) = 1$.

3.8 Multivariate Distributions and Covariance

- Distribution of multiple variables:
- Elements belong to real vector space $\mathbb{R}^n$.
- $P(AB) \ldots P(A, B)$
- $p(x_1, x_2, \ldots, x_n) \geq 0$ is the joint probability distribution function (PDF).
- $\int_\Omega p(\vec{x}) d^n x = 1$
- $E(f(\vec{x})) = \int_\Omega f(\vec{x}) p(\vec{x}) d^n x$
- $\mu_i = \int x_i p(\vec{x}) d^n x$
- $V(x_i) = \sigma_i^2 = \int (x_i - \mu_i)^2 p(\vec{x}) d^n x$
- Covariance:
- $V_{i,j} = E((x_i - \mu_i)(x_j - \mu_j))$
- $V_{i,i} = \sigma_i^2 = E((x_i - \mu_i)^2)$ (variance)
- $V_{i,j} = V_{j,i}$ (symmetry)

4 Joint Distributions, Correlations, and Variable Transformations

Thursday, September 18th 2025

4.1 Conditional Probability: A Simple Example

- For fun, example that depends on cultural assumptions: A king comes from a family with two kids. What is the probability that the king's sibling is a sister?
- $S = \{(m,m),(m,f),(f,m),(f,f)\}$
- $P(S|K) = \frac{P(SK)}{P(K)} = \frac{\frac{1}{2}}{\frac{3}{4}} = \frac{2}{3}$

4.2 Distributions of Multiple Random Variables

- $p(x_1, x_2, \ldots, x_n)$
- $S = \mathbb{R}^n$
- $\int p(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2\, \ldots\, dx_n = 1$
- For any function $f(\vec{x})$:
$$E(f) = \int f(\vec{x}) p(\vec{x}) d\vec{x}$$
- $E(x_1) = \int x_1 p(\vec{x}) d\vec{x} = \mu_1$
- $E(x_i) = \mu_i$
- $V(x_i) \equiv \sigma_i^2 = \int (x_i - \mu_i)^2 p(\vec{x}) d\vec{x}$

4.3 Covariance Matrix and Correlation Coefficient

- Define covariance:
$$V_{ij} = E((x_i - \mu_i)(x_j - \mu_j))$$
- $V_{ii} = \sigma_i^2$ (variance)
- $V_{ij} = V_{ji}$ (symmetry)
- $V_{ij} = 0$ for independent variables
- Expanding the covariance matrix:
$$\begin{aligned}
V_{ij}(\vec{x}) &= E((x_i - \mu_i)(x_j - \mu_j)) \\
&= E(x_i x_j - \mu_i x_j - \mu_j x_i + \mu_i \mu_j) \\
&= E(x_i x_j) - \mu_i E(x_j) - \mu_j E(x_i) + \mu_i \mu_j \\
&= E(x_i x_j) - \mu_i \mu_j - \mu_j \mu_i + \mu_i \mu_j \\
&= E(x_i x_j) - \mu_i \mu_j
\end{aligned}$$
- So we can say that $V_{ij} \geq 0$
- V_{ij} can be negative, zero, or positive

- Define the correlation coefficient:

$$\rho(x_i, x_j) = \rho_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}}\sqrt{V_{jj}}} = \frac{V_{ij}}{\sigma_i \sigma_j}$$

- We find that $-1 \leq \rho_{ij} \leq 1$

4.4 Independence and Uncorrelated Variables

- Random variables $x_i, \ldots, x_n$ are independent if the joint pdf factorizes:

$$p(x_1, \ldots, x_n) = p_1(x_1)p_2(x_2)\ldots p_n(x_n)$$

- Independent variables are uncorrelated:

$$\begin{aligned} E(x_i x_j) &= \int x_i x_j p(\vec{x}) d\vec{x} \\ &= \int x_i x_j p_1(x_1)\ldots p_n(x_n) dx_1 \ldots dx_n \\ &= \int x_i p_i(x_i) dx_i \int x_j p_j(x_j) dx_j \int p_2(x_2) dx_2 \ldots \int p_n(x_n) dx_n = \mu_i \mu_j \end{aligned}$$

$$V_{ij} = E(x_i x_j) - \mu_i \mu_j$$

In the case of independent variables:

$$V_{ij} = \mu_i \mu_j - \mu_i \mu_j = 0$$

- Independent variables are uncorrelated, but uncorrelated variables are not necessarily independent.

4.5 Examples of Correlated and Uncorrelated Variables

- 100% correlation example:
- $x = \text{Uniform}[-1, 1]$, plot distribution from -1 to 1.
- $y = x$:
- $V_{ij} = E(xy) - E(x)E(y) = E(x^2) = \int_{-1}^{1} x^2 \frac{1}{2} dx = \frac{1}{3} \neq 0$
- $y = |x|$:
- $E(xy) = \int_{-1}^{0} x(-x)p(x)dx + \int_0^1 xxp(x)dx$
- $E(xy) = \int_0^1 x^2 \frac{1}{2} dx - \int_{-1}^{0} x^2 \frac{1}{2} dx = \frac{1}{6} - \frac{1}{6} = 0$

4.6 Marginal Distributions

- For a joint pdf $p(x_1, x_2, \ldots, x_n)$, the marginal probability density functions are:

$$f_1(x_1) = \int p(x_1, x_2, \ldots, x_n) dx_2 dx_3 \ldots dx_n$$

- If variables are independent:

$$\begin{aligned} f_1(x_1) &= \int p(x_1, x_2, \ldots, x_n) dx_2 \ldots dx_n \\ &= p_1(x_1) \int p_2(x_2) dx_2 \int p_3(x_3) dx_3 \ldots \int p_n(x_n) dx_n \\ &= p_1(x_1) \cdot 1 \cdot 1 \cdot \ldots \cdot 1 = p_1(x_1) \end{aligned}$$

4.7 Change of Variables in Probability Densities

- Something we need to know, because we do it all the time:
  - Change of variables of P
  - Calculate new V_{ij} under new variables
- Let x be a random variable with pdf $f(x)$ and let y be some function.
- First: y is one-to-one with f

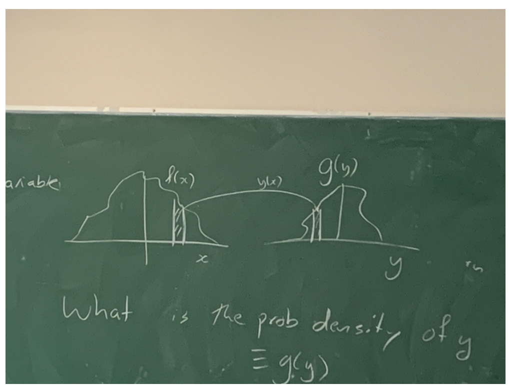

Figure 4.1: 1-to-1 function

- What is the probability density of y, denoted $g(y)$?
- Conservation of probability:
- $f(x)dx = g(y)dy$
- $g(y) = f(x) \left| \frac{dx}{dy} \right|$

$$\boxed{f(x) \left| \frac{dx}{dy} \right| = g(y)}$$

4.8 Change of Variables: Non One-to-One Case

- If y is not one-to-one: sum over all segments that map to the same y.
- Example: $f(x)$ uniform on $[0,1]$, $f(x) = 1$
- Let $y(x) = \frac{-1}{\lambda}\ln(x)$
- $\frac{dy}{dx} = \frac{-1}{\lambda x}$
- $\frac{dx}{dy} = -\lambda x$
- $-\lambda x = \ln x$
- $e^{-\lambda y} = x$
- $\lambda > 0 \Rightarrow \frac{dx}{dy} = -\lambda x = -\lambda e^{-\lambda y}$
- $g(y) = f(x)\left|\frac{dx}{dy}\right| = 1 \cdot \lambda e^{-\lambda y} = \lambda e^{-\lambda y}$

4.9 Multivariate Transformations and the Jacobian

- If we have variables $\{x_i\}$ and transform to new variables $\{y_i\}$:
- Region $\mathbb{R}$ in x-space maps to region $\mathbb{R}'$ in y-space.

$$\int_{\mathbb{R}} f(\vec{x})d\vec{x} = \int_{\mathbb{R}'} f(\vec{x}(\vec{y}))(\vec{y})\left|\frac{\partial \vec{x}}{\partial \vec{y}}\right| d\vec{y}$$

$$g(\vec{y}) = f(\vec{x}(\vec{y}))\,|J|$$

- Where $\left|\frac{\partial \vec{x}}{\partial \vec{y}}\right|$ is the Jacobian determinant of the transformation.
- Jacobian matrix J:

$$J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}$$

4.10 Example: Cartesian to Polar Transformation

- Change to polar coordinates:
- $x = r\cos\theta$
- $y = r\sin\theta$
- $P'(r,\theta) =? = p(x,y)\left|\frac{\partial(x,y)}{\partial(r,\theta)}\right|$
- $\frac{\partial x}{\partial r} = \cos\theta$
- $\frac{\partial y}{\partial r} = \sin\theta$

- $\frac{\partial x}{\partial \theta} = -r \sin \theta$
- $\frac{\partial y}{\partial \theta} = r \cos \theta$
- $J = \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{bmatrix}$
- $J = r \cos^2 \theta + r \sin^2 \theta = r$
- $p'(r, \theta) = \frac{r}{\pi} dr d\theta$

5 Propagation of Uncertainty in Measurements

Tuesday, September 20th 2025

5.1 Propagation of Errors for a Single Variable

- Given $f(x)$ pdf, $\mu \equiv E(x)$, $\sigma^2 \equiv V(x) = E(x^2) - \mu^2$
- Know $f(x) \to g(y)$, given $y(x)$.
- Taylor expand $y(x)$ about mean μ:

$$y(x) = y(\mu) + y'(\mu)(x-\mu) + \frac{1}{2!}y''(\mu)(x-\mu)^2 + \dots$$

$$E(y(x))? \equiv \mu_y$$

$$E(y(x)) = E(y(\mu)) + y'(\mu)E(x-\mu) + \frac{1}{2!}y''(\mu)E((x-\mu)^2) + \dots$$

$$= y(\mu) + y'(\mu)\cdot 0 + \frac{1}{2!}y''(\mu)V(x) + \dots$$

- To the 1st order:

$$\mu_y = E(y(x)) = y(\mu) = y(E(x))$$

5.2 Variance Propagation for a Single Variable

- Variance of y:

$$\begin{aligned}
V(y) &= E((y(x) - E(y(x)))^2) && (5.1)\\
&= E((y(x) - \mu_y)^2) && (5.2)\\
&= E((y'(\mu)(x-\mu) + \frac{1}{2!}y''(\mu)(x-\mu)^2 + \dots)^2) && (5.3)\\
&= E(y'(\mu)^2(x-\mu)^2 + y'(\mu)y''(\mu)(x-\mu)^3 + O((x-\mu)^4)) && (5.4)\\
&= y'(\mu)^2 V(x) + \dots && (5.5)
\end{aligned}$$

- Some relations:

$$\begin{aligned}
E(x) &\equiv \mu_x\\
V(x) &\equiv \sigma_x^2\\
y &= y(x)\\
E(y) &\equiv \mu_y = y(\mu_x)\\
V(y) &\equiv \sigma_y^2 = (y'(\mu_x))^2\sigma_x^2
\end{aligned}$$

$$\sigma_y = |y'(\mu_x)|\sigma_x$$

- Example: $y = \frac{1}{x}$, $\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{1}{x^2}$

$$\sigma_y^2 = \frac{1}{\mu_x^4}\sigma_x^2$$

5.3 Propagation of Errors for Multiple Variables

- Let us suppose we have n variables $\{x_i\}$, with pdf $f(\vec{x})$.
- Let $y_j = 1, 2, \ldots, m$ be m functions of x_i.
- $y_j = y_j(x_1, x_2, \ldots, x_n)$
- $V_{ij}(x)_{n\times n}(\vec{x})$ = covariance matrix of $\{x_i\}$
- $V_{ij}(\vec{x}) = E((x_i - \mu_i)(x_j - \mu_j))$
- Taylor expand each y_j: $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$
- $y_j(\vec{x}) = y_j(\vec{\mu}) + \sum_i \frac{\partial y_j}{\partial x_i}\Big|_{\vec{\mu}} (x_i - \mu_i) + \frac{1}{2!}\sum_{i,k} \frac{\partial^2 y_j}{\partial x_i \partial x_k}\Big|_{\vec{\mu}} (x_i - \mu_i)(x_k - \mu_k) + \ldots$
- $E(y_j(\vec{x})) = E(y_j(\vec{\mu})) + \sum \frac{\partial y_j}{\partial x_i} E(x_i - \mu_i) + \ldots = y_j(\vec{\mu})$

5.4 Covariance Propagation for Functions of Multiple Variables

- Covariance between y_k and y_l:

$$\begin{aligned} &E((y_k - \mu_{y_k})(y_l - \mu_{y_l})) \\ &= E((y_k - y_k(\mu))(y_l - y_l(\mu))) \\ &= E\left(\sum_i \frac{\partial y_k}{\partial x_i}\Big|_\mu (x_i - \mu_i) \sum_j \frac{\partial y_l}{\partial x_j}\Big|_\mu (x_j - \mu_j)\right) \\ &= \sum_{i,j} \frac{\partial y_k}{\partial x_i}\Big|_\mu \frac{\partial y_l}{\partial x_j}\Big|_\mu E((x_i - \mu_i)(x_j - \mu_j)) \end{aligned}$$

$$\boxed{V_{kl}(\vec{y})_{m\times m} = \sum_{i,j} \frac{\partial y_k}{\partial x_i}\Big|_{\vec{\mu}} \frac{\partial y_l}{\partial x_j}\Big|_{\vec{\mu}} V_{ij}(\vec{x})_{n\times n}}$$

- Example: x, y random variables,

$$V(x,y) = \begin{bmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

- $z = x + y$
- $V(z) = \sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 V_{xx} + 2\frac{\partial z}{\partial x}\frac{\partial z}{\partial y} V_{xy} + \left(\frac{\partial z}{\partial y}\right)^2 V_{yy}$
- $= \sigma_x^2 + 2\rho_{xy}\sigma_x\sigma_y + \sigma_y^2$
- If x_i are uncorrelated,

$$V_{ij} = \sigma_{i,j}\sigma_i^2 = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$V_{kl}(\vec{y}) = \sum_i \frac{\partial y_k}{\partial x_i}\Big|_\mu \frac{\partial y_l}{\partial x_i}\Big|_\mu V_{ii}(\vec{x})$$

$$\text{variance } V_{kk} = \sum_i \left(\frac{\partial y_k}{\partial x_i}\right)^2 \sigma_i^2$$

5.5 Examples of Error Propagation in Measurements

- Example: Measuring resistances. x_i independent, $z = x + y$, $x = R_1$ resistor value, $y = R_2$ resistor value, $z = R_{\text{tot}}$ total resistance.
- $R_1 \pm \sigma_{R_1}$
- Convention is to use $\sqrt{V(R)}$ as uncertainty.
- For a good measuring device, $E(R) = R_{\text{true}} \leftarrow$ unbiased.
- $V(R) =$ small
- $R_1 \pm \sigma_{R_1}$, $R_2 \pm \sigma_{R_2}$, then $\sigma_{R_{\text{tot}}} = \sqrt{\sigma_{R_1}^2 + \sigma_{R_2}^2}$
- $R = R_{\text{tot}} = R_1 + R_2$
- $z = xy$, like I, R
- $\sigma_z^2 = \left(\frac{\partial z}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial z}{\partial y}\right)^2 \sigma_y^2 = y^2\sigma_x^2 + x^2\sigma_y^2$

$$\boxed{\left(\frac{\sigma_z}{z}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2}$$

5.6 Matrix Formulation of Linear Error Propagation

- Formula is exact if transformation of variables is linear.
- $\vec{y} = A\vec{x}$, A is $m \times n$ matrix, $\vec{x}$ is $n \times 1$, $\vec{y}$ is $m \times 1$.
- $\frac{\partial y_k}{\partial x_i} =$ constant $\Rightarrow$ higher order terms in Taylor expansion are 0
- $V_{kl}(\vec{y}) = \sum_{i,j} \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} V_{ij}(\vec{x})$
- Matrix notation:
- $V_{kl}(\vec{y}) = \sum_{i,j} A_{ki} A_{lj} V_{ij}(\vec{x})$
- $= \sum_{i,j} A_{ki} V_{ij}(\vec{x}) A_{lj}$
- $= \sum_{i,j} A_{ki} V_{ij} (A^T)_{jl}$
- $= (AV(\vec{x})A^T)_{kl}$

$$\boxed{V(\vec{y})_{m\times m} = A_{m\times n} V(\vec{x})_{n\times n} A^T_{n\times m}}$$

5.7 Variance of the Arithmetic Mean

- Example: Arithmetic mean. Let $x_i = n$ identical independent variables with $V(x_i) = \sigma_x^2$
- Set $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$
- Recall that $V(ax) = a^2 V(x)$
- $V(\bar{x}) = V\left(\frac{1}{n}\sum_{i=1}^n x_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n x_i\right) = \frac{1}{n^2}\sum_{i=1}^n V(x_i) = \frac{1}{n^2} n\sigma_x^2 = \frac{\sigma_x^2}{n}$

- If variables are different σ_i^2: n measurements
- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
- $V(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^{n} V(x_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2$
- $\sigma_{\bar{x}} = \frac{1}{n} \sqrt{\sum_{i=1}^{n} \sigma_i^2}$

5.8 Example: Measuring the Period of a Sine Wave

- Example: Measure period of sine wave on scope.
- $T = \Delta t = t_2 - t_1$
- $\sigma_T^2 = \left(\frac{\partial \Delta t}{\partial t_1}\right)^2 \sigma_t^2 + \left(\frac{\partial \Delta t}{\partial t_2}\right)^2 \sigma_t^2 = \sigma_t^2 + \sigma_t^2 = 2\sigma_t^2$
- Measure N cycles, $T = \frac{1}{N} \Delta t$
- $\sigma_{T^2} = \frac{1}{N^2} \sigma_{\Delta t}^2 = \frac{2}{N^2} \sigma_t^2$

6 Covariance Transformations and the Binomial Distribution

Tuesday, September 25th 2025

6.1 Covariance Transformation Under Linear Transformations

- Linear transformation:
$$\vec{y} = A\vec{x}$$
$$V_{kl}(\vec{y}) = \sum_{i,j} \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} V_{ij}(\vec{x})$$
- Linear $y_k = \sum A_{kj} x_j$
- then
$$V_{kl}(\vec{y}) = \sum_{i,j} A_{ki} A_{lj} V_{ij}(\vec{x})$$
- or in matrix form
$$V(\vec{y}) = \left(AV(\vec{x})A^T\right)_{kl}$$

Diagonalization via Eigenvectors

- If $\hat{e}_i$ are the eigenvectors of V, then
$$V(\vec{x})\hat{e}_i = \lambda_i \hat{e}_i$$
- Form:
$$A = \begin{pmatrix} \hat{e}_1 \\ \dots \\ \hat{e}_n \end{pmatrix} = \begin{pmatrix} \hat{e}_{11} & \hat{e}_{12} & \dots & \hat{e}_{1n} \\ \dots & \dots & \dots & \dots \\ \hat{e}_{n1} & \hat{e}_{n2} & \dots & \hat{e}_{nn} \end{pmatrix}$$
- then
$$A^T A = I$$
- then:
$$VA^T = V \begin{pmatrix} \hat{e}_1 & \dots \\ \dots & \dots \\ \hat{e}_n & \dots \end{pmatrix} = \begin{pmatrix} \lambda_1 \hat{e}_{11} & \dots & \lambda_n \hat{e}_{n1} \\ \dots & \dots & \dots \\ \lambda_1 \hat{e}_{1n} & \dots & \lambda_n \hat{e}_{nn} \end{pmatrix}$$
- Then:
$$AVA^T = \begin{pmatrix} \hat{e}_{11} & \dots & \hat{e}_{1n} \\ \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \lambda_1 \hat{e}_{11} & \dots & \\ \dots & \dots & \dots \\ \lambda_1 \hat{e}_{1n} & \dots & \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$
- Then:
$$AVA^T = V(\vec{y}) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

6.2 The Binomial Distribution

- Consider an experiment with two outcomes.
- E.g. coin flips, selecting a ball with 2 possible colours, etc.
- One trial is called a Bernoulli trial.

Bernoulli Trials and Sampling Methods

- Example – Method 1: You have an urn filled with N balls. Some are red (R), some are blue (B).
- (0) What is your estimate of n_R, n_B, or $f = n_R/N$ or p of drawing R?
- (1) You pick a ball: R. Q: estimate of $p = n_R/N$?
- (2) You pick another without replacing 1st ball: get R.
- (3) R
- (4) Get B
- This is a question about this ONE urn.
- Now Method 2: you draw red, and you PUT IT BACK. You repeat this several times.
- Now Method 3: We have an infinite source of balls with fraction p red and $(1-p)$ blue.

$$P(R) = p$$
$$P(B) = 1 - p$$

Derivation of the Binomial Probability

- Make infinite number of urns all with N balls, with fraction p red and $(1-p)$ blue.
- Open all, count n_R red balls, n_B blue balls.
- In our case we have N balls, prob $p = R$ and $1 - p = q = B$.
- Prob of getting sequence RRB is:

$$P(RRB) = p \cdot p \cdot (1-p) = p^2(1-p)$$

- If we don't care about order, then:

$$P(RRB) = P(RBR) = P(BRR) = p^2(1-p)$$

- There are 3 ways of ordering RRB, so total probability is:

$$P(2R, 1B) = 3p^2(1-p) = 3p^2q$$

- Number of ways to choose r items from N is:

$$\binom{N}{r} = \frac{N!}{r!(N-r)!}$$

- Probability of getting exactly r R out of N:

$$P_r = \binom{N}{r} p^r (1-p)^{N-r} = B(r; N, p)$$

- This is called the Binomial distribution and applies to anything where there are 2 outcomes $(A, \bar{A})$.

Mean and Variance of the Binomial Distribution

- Want mean, σ

$$E(r) = \sum_{r=0}^{n} rP_r = \sum_{r=0}^{n} r\binom{n}{r} p^r (1-p)^{n-r}$$
$$= \sum_{r=0}^{n} r\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}$$
$$= \sum_{r=1}^{n} \frac{n!}{(r-1)!(n-r)!} p^r (1-p)^{n-r}$$
$$= np\sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} (1-p)^{n-r}$$

- Change sum $r' = r-1 \to n' = n-1$

$$E(r) = np\sum_{r'=0}^{n-1} \frac{(n-1)!}{r'!(n-1-r')!} p^{r'} (1-p)^{(n-1)-r'}$$
$$= np\sum_{r'=0}^{n-1} \binom{n-1}{r'} p^{r'} (1-p)^{n'-r'}$$
$$= np \cdot 1 = np$$

- from:

$$(p+q)^n = \sum_{r=0}^{n}\binom{n}{r} p^r q^{n-r}$$
$$(p+1-q)^n = 1^n = 1$$
$$E(r) = np$$

- This is what we want!
- Now:

$$V(r) = \sum r^2 p_r - E(r)^2 = \sum r^2 p_r - n^2p^2$$

- Slightly easier to calculate:

$$\sum r(r-1)p_r = \sum_{r=0}^{n} r(r-1)\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r} = \sum_{r=2}^{n} \frac{n!}{(r-2)!(n-r)!} p^r q^{n-r}$$
$$= n(n-1)p^2 \sum_{r=2}^{n} \frac{(n-2)!}{(r-2)!(n-r)!} p^{r-2} q^{n-r}$$

- Sub $r' = r-2$

$$= n(n-1)p^2 \sum_{r'=0}^{n-2} \binom{n-2}{r'} p^{r'} q^{(n-2)-r'}$$
$$= n(n-1)p^2 \cdot 1 = n(n-1)p^2$$
$$= n^2p^2 - np^2$$

- Such that:

$$\boxed{V(r) = \sum r^2 p_r - n^2p^2} = np(1-p) = npq$$

Applications to Histograms and Counting Statistics

- Why is this important? Histograms are often Binomially distributed.
- Data either falls A: falls in bin, or $\bar{A}$: does not fall in bin.
- $p =$ probability of falling in ith bin.
- $+n$ entries, e.g. students in class, histogram = grades.
- Expected number of entries is np.
- Plot of taking distribution several times and checking how many fall in bin i and then plotting that distribution is Binomial.
- Usually you have 1 histogram.
- Look at entries in bin i – n_i/n = fraction of entries in bin i.
- Estimator $p = n_i/n$.
- Expect if you repeated $\Rightarrow n_i$ would follow Binomial distribution with mean $\sigma_i = \sqrt{npq}$ and $V_i = \sigma_i^2 = np(1-p)$.
- $$p = \frac{n_i}{n}$$
- $$\boxed{\sigma_i = \sqrt{n_i\left(1 - \frac{n_i}{n}\right)}} \approx \sqrt{n_i} \text{ if } n \gg n_i$$
- Notes: we do know the total number n, how often is it in bin i.
- HW: given the distribution, how many times n do I need to do it to get that.
- r fixed n, vs. n fixed r.