§2

Fundamentals of probability theory.

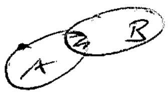
<u>Example 1</u> fair coin: $p_{head} = p_{tail} = \frac{1}{2}$.

<u>Axioms</u> of probability

$X \equiv$ set of states.

$0 \le p(A) \le 1$ for (some) $A \subseteq X$

$p(\emptyset) = 0, \quad p(X) = 1.$



$p(A) + p(B) = p(A \cup B) - p(A \cap B)$

<u>Example 2</u> (Continuous state space)

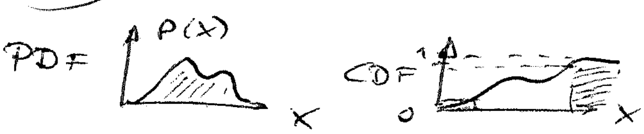
$X = \mathbb{R}$ (e.g. one-dim'l position)

$p(x) : \mathbb{R} \to \mathbb{R} \equiv$ probability density

$P(A) = \int_A p(x)\,dx.$

N.B. If x has physical units of length (m), $p(x)$ has units of one-dim'l density (1/m)

N.B.



PDF $\quad p(x) \equiv$ probability density function

CDF $C(x) = \int_{-\infty}^{x} dx'\, p(x') \equiv$ cumulative ~~prob.~~ density function

percentiles: $\alpha = C(x_\alpha)$

Probabilities in Physics

- probability = relative frequency

$$p(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

$\downarrow$ requires

a) experiment that can be repeated with same initial macro-state.

b) stochastic simulations as idealized model of physical reality.

<u>Example</u> (weather forecast)

R = rain tomorrow (≥ 1 mm)

frequentist statistics.

$$p(R) = \frac{116}{365} = 31.8\%$$

$$p(R \mid \text{October}) = \frac{8}{31} = 25.8\%.$$

persistence ($\rightarrow$ Markov models).

$p(R \mid \text{current local weather}) = x$

mathematical forecasting

$p(R \mid \text{current global weather with measurement uncertainty}) = y$

$p(R \mid \text{rain today}) = 44\%$, $p(R \mid \text{dry today}) = 17\%$ Castey [illegible] 1983

Moments

$$\mu_n = \langle x^n \rangle = \int_{-\infty}^{\infty} dx\, x^n \cdot p(x).$$

N.B characteristic function

$$\langle \exp tx \rangle = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!}$$

Moments are Taylor coefficients of characteristic function.

Cumulants

$$k_1 = \mu_1 \equiv \text{mean}$$

$$k_2 = \mu_2 - \mu_1^2 \equiv \text{variance}.$$

$$k_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3.$$

More generally

$$\ln \langle \exp tx \rangle = \sum_{n=0}^{\infty} k_n \frac{t^n}{n!}$$

Example (normal distribution)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] = \mathcal{N}(\mu, \sigma^2)$$

$$\mu_1 = \mu, \quad \mu_2 = \mu^2 + \sigma^2, \quad \mu_3 = \mu^3 + 3\mu\sigma^2, \ldots$$

$$k_1 = \mu, \quad k_2 = \sigma^2, \quad \underline{k_3 = 0 \quad n \geq 3} \ !$$

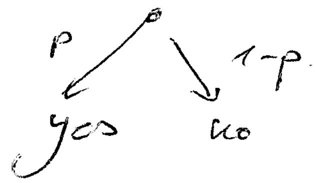
all higher cumulants are zero.

Important distributions

1. Normal distribution

2. Binomial distribution

- Bernoulli trial



- n independent trials

$$p(k,n) = p(k \times \text{"yes"}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\equiv$ Bernoulli distribution

- $\langle k \rangle = np$.

$$\langle k^2 \rangle - \langle k \rangle^2 = np(1-p)$$

Proof: $\langle k \rangle = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$
$= \sum_{k=1}^{n}$ —"— $= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np$.

- Normal approximation

$$p(k,n) \approx \mathcal{N}(np,\ np(1-p))$$

$\rightarrow$ later.

Exercise: proof.

Hint: use Stirling formula.

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

Poisson distribution

We consider a "continuous-time limit" of the Bernoulli distribution

- introduce time $t_j = \frac{j}{n} T$.
- $\lambda = pn$ expected total number of events
- $r = \frac{\lambda}{T} = \frac{p}{dt} \equiv$ event rate $[\frac{1}{s}]$
- limit $n \to \infty$ while $\lambda = $ const

$\Rightarrow$ e.g. $p = \lambda/n$.

$$p(k,\lambda) = \exp(-\lambda) \cdot \frac{\lambda^k}{k!}$$

$\equiv$ Poisson distribution
k events (indep. with rate $r = \frac{\lambda}{T}$) in time int. T.

$\mu = \langle k \rangle = \lambda$

$\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2 = \lambda$

Exercise: Check $\sum p(k) = 1$

Exercise*: Derive $p(k)$, μ, σ^2 from Bernoulli distribution, using Stirling formula

Normal approximation

$$p(k,\lambda) \approx \mathcal{N}(\lambda, \lambda) \quad \text{for } \lambda \text{ large}$$

Examples

- customers arriving at counter, service requests arriving at computer server
- radioactive decay events.
- photons arriving at CCD-sensor (shot noise)

Power-law distributions

$$p(x) \sim \frac{1}{x^{\alpha}} \quad \text{for} \quad x \gg 1$$

$\equiv$ "fat tail"

- have unpleasant properties
- variance $= \infty$ for $\alpha < 3$.

Examples

- jump distribution of certain foraging strategies (Levy flights)
- interaction networks (e.g. facebook contacts, $\alpha = 2.2$).

Reminder: Distribution for single random var.

- Binomial d.: n indep. "yes/no" trials
- Poisson d.: number of events with rate r in fixed time interval
- Normal distribution.

Today: - conv. to normal distrib.
- Stoch. processes: family of random variables

Normal approximation of Bernoulli distribution

$$p(k,n) = \binom{n}{k} p^{k} (1-p)^{n-k}.$$

$$q = 1-p.$$

$$k = np + n\varepsilon \Rightarrow p(k,n) \neq 0 \text{ for } \varepsilon = O\left(\tfrac{1}{\sqrt{n}}\right)$$
$$\approx 0 \text{ for } \varepsilon \gg \tfrac{1}{\sqrt{n}}$$

- Trick 1: Stirlings formula.

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^{n}$$

$$p(k,n) = \underbrace{\frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}}}_{\frac{1}{\sqrt{2\pi npq}} + O(\varepsilon)} \underbrace{\frac{n^{n}}{k^{k}(n-k)^{n-k}} \cdot p^{k} q^{n-k}}_{\left(\frac{np}{k}\right)^{k}\left(\frac{nq}{n-k}\right)^{n-k}}.$$

- Trick 2: $x^{k} = e^{k \cdot \ln x}$

- $\ln\left(\frac{np}{k}\right) = \ln\left(\frac{p}{p+\varepsilon}\right) = -\ln\left(1+\frac{\varepsilon}{p}\right).$

$$\approx -\frac{\varepsilon}{p} + \frac{1}{2}\left(\frac{\varepsilon}{p}\right)^{2} - \ldots$$

- $\ln\left(\frac{nq}{n-k}\right) = \ln\left(\frac{q}{q-\varepsilon}\right) = -\ln\left(1-\frac{\varepsilon}{q}\right)$

$$\approx +\frac{\varepsilon}{q} - \frac{1}{2}\left(\frac{\varepsilon}{q}\right)^{2} + \ldots$$

$$\Rightarrow \left(\frac{np}{k}\right)^{k}\left(\frac{nq}{n-k}\right)^{n-k} \approx \exp\left(\underbrace{k\left[-\frac{\varepsilon}{p} + \frac{1}{2}\left(\frac{\varepsilon}{p}\right)^{2}\right] + (n-k)\left[\frac{\varepsilon}{q} - \frac{1}{2}\left(\frac{\varepsilon}{q}\right)^{2}\right]}\right)$$

$$n(p+\varepsilon)\left[-\frac{\varepsilon}{p} + \frac{1}{2}\left(\frac{\varepsilon}{p}\right)^{2}\right] + n(q-\varepsilon)\left[\frac{\varepsilon}{q} - \frac{1}{2}\left(\frac{\varepsilon}{q}\right)^{2}\right]$$

$$= 0 \cdot \varepsilon - \frac{1}{2} n \frac{\varepsilon^{2}}{p} - \frac{1}{2} n \frac{\varepsilon^{2}}{q} + O(\varepsilon^{3})$$

$$= -\frac{1}{2}\frac{n\varepsilon^2}{pq}\underbrace{(p+q)}_{1} = -\frac{1}{2}\frac{(k-np)^2}{npq}.$$

$\Longrightarrow$

$$p(k,n) \simeq \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(k-np)^2}{2\sigma^2}$$

$$\sigma^2 = npq.$$

<u>Second proof</u> (Central limit theorem)

indep. random variables (n Bernoulli trials)

$$X_j = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{— " —} \quad 1-p \end{cases}, \quad j = 1, \dots, n$$

$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) \equiv$ empirical mean

$k = n \cdot \bar{X}$.

$p(k,n) = p(\bar{X}) \sim \mathcal{N}(np, npq)$.

(by Central-limit theorem later today)

Central-limit theorem.

$X_1, \ldots, X_n \equiv$ independent, identically distributed random variables, mean μ, variance σ^2.

$$\overline{X} = \frac{X_1 + \ldots + X_n}{n} \equiv \text{mean}$$

$\longrightarrow$ itself a random variable.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}.$$

Theorem

$$p(Z) \xrightarrow{D} \mathcal{N}(0,1) \text{ for } n \text{ large}$$

$\equiv$ "convergence in distribution"

or equivalently

$$CDF(Z) \longrightarrow \mathrm{Erf}(Z) \text{ for almost all } Z \in \mathbb{R}.$$

Idea of proof:

Cumulants of X_i: $k_1 = \mu, k_2 = \sigma^2, k_3, k_4, \ldots$

— " — of $\overline{X}$, $k_1 = \mu, k_2 = \frac{\sigma^2}{n}, k_3 \sim \frac{1}{n^2}, k_4 \sim \frac{1}{n^3}$

$*$ — " — of Z: $k_1 = 0, k_2 = 1, k_3 \sim \frac{1}{\sqrt{n}}, k_4 \sim \frac{1}{n}, \ldots$

$\Rightarrow$

$$\lim_{n\to\infty} \langle \exp iZ \rangle = 1 - \frac{\sigma^2 t^2}{2}$$

$\Rightarrow \ldots \Rightarrow \qquad Z \xrightarrow{D} \mathcal{N}(0,1)$. q.e.d.

- $C_X(t) = \langle \exp tX \rangle \equiv$ characteristic function

$$C_{\alpha X}(t) = C_X(\alpha t), \quad \alpha \in \mathbb{R}.$$

$$\langle (\alpha X)^j \rangle = \alpha^j \langle X^j \rangle$$

$$k_{\alpha X, j} = \alpha^j k_{X,j} \text{ for cumulants.}$$

Use ~~this~~ for $\bar{X}$, Z.

$$C_{\bar{X}}(t) = C_X\left(\frac{t}{n}\right)^n$$

$$C_Z(t) = C_{\bar{X}}\left(\frac{t}{\sigma/\sqrt{n}}\right) \exp\left(-\frac{\mu t}{\sigma/\sqrt{n}}\right)$$

$$= C_X\left(\frac{t}{\sigma\sqrt{n}}\right)^n \exp\left(-\frac{\mu t}{\sigma/\sqrt{n}}\right)$$

$$\ln C_Z(t) = n \ln C_X\left(\frac{t}{\sigma\sqrt{n}}\right) - \frac{\mu t}{\sigma/\sqrt{n}}.$$

$$k_{Z,j} = n\left(\frac{1}{\sigma\sqrt{n}}\right)^j k_{X,j} \quad \text{for } j \geq 2.$$

i.e.

$$k_{Z,2} = \underline{1}$$

$$k_{Z,3} \sim \frac{1}{\sqrt{n}}$$

$$k_{Z,4} \sim \frac{1}{n} \quad \ldots$$

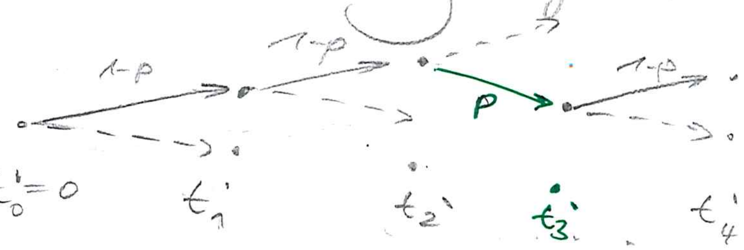
Stochastic processes

- (generalized) random function $f(t): \mathbb{R} \to \mathbb{R}$
  $\equiv$ family of random variables $f(t)$, parametrized by t.

Example 1: Poisson point process:

$$f(t) = \sum_{i=-\infty}^{\infty} \delta(t - t_i)$$

$t_i \equiv$ random event times, of events occuring independently with rate r.

- The Poisson point process can be constructed as 'time-continuous limit' of the Bernoulli distribution
  - $dt \equiv$ discrete time-step.
  - $t_j' = j \cdot dt$, $j \in \mathbb{Z}$ $\equiv$ discrete time
  - $p = r \cdot dt$ $\equiv$ probability for event during single time step.
  - 

  - if event at t_i', set next $t_j := t_i'$
- Relation to Poisson distribution

$T \equiv$ observation time

$N = \int_0^T dt\, f(t) \equiv$ number of events in $[0,T]$.

$p(N = k) = p(k, \lambda = rT)$.

$\langle f(t) \rangle = r$, $\langle N \rangle = \lambda$.

Terminology

- Conditional probability density.

$$p(f(t_2)=f_2 \mid f(t_1)=f_1)$$

[sometimes: $p(f_2, t_2 \mid f_1, t_1)$]

- Markov property

For $t_3 > t_2 > t_1$

$$p(f(t_3)=f_3 \mid f(t_2)=f_2, f(t_1)=f_1)$$
$$= p(f(t_3)=f_3 \mid f(t_2)=f_2)$$
$$\forall t_j, f_j$$

Example: diffusion.
Counter-example: drawing from an urn without replacement.

- Martingale

$$\langle f(t_2)=f_2 \mid f(t_1)=f_1 \rangle = f_1 \quad \forall t_j, f_j$$

Counterexample: diffusion with drift.

Examples

- Bernoulli process (time. discrete).
- Poisson process.
- Wiener process.

[Poisson point process cont.]

Waiting times are exponentially distributed.

$$t = t_{i+1} - t_i \equiv \text{inter-event times}$$

$$p(t) = r\exp(-rt)$$

Proof

CDF: $P(t \geq \theta + dt) =$

$$P(t \geq \theta) - r\,dt\, P(t \geq \theta) + O(r\,dt)^2$$

$$\Rightarrow \frac{d}{dt} P(t \geq \theta) = -r\, P(t \geq \theta)$$

$$\Rightarrow P(t \geq \theta) \sim \exp(-rt) \qquad \text{q.e.d.}$$

N.B. $r\,dt$ dimensionless, thus $r\,dt \ll 1$ meaningful.

Example 2: Gaussian white noise

$\xi(t): \mathbb{R} \to \mathbb{R}$. such that.

(i) $\langle \xi(t) \rangle = 0$, (ii) $\langle \xi(t)\, \xi(t') \rangle = 2D\, \delta(t-t')$.

(iii) $\int_{t_1}^{t_2} dt \;\; \xi(t) \sim N(0, 2D\, |t_2 - t_1|)$.

N.B. Gaussian noise because of (iii)

N.B. white noise because of

$$S_\xi(\omega) = \langle \tilde{\xi}(\omega)\, \tilde{\xi}(\omega')^* \rangle = 2D\, \delta(t-t').$$

$\equiv$ flat power spectral density.

N.B. Mathematicians define

$$W(t) = \int_0^t dt \;\; \xi(t) \equiv \text{Wiener process},$$

which is continuous (with prob. 1) and can be shown to exist in a rigorous sense.

Exercise: $W(t_2) - W(t_1) \sim N(0, 2D\, |t_2 - t_1|)$.

$W(0) = 0$.

$W(t_2) - W(t_1)$ indep. of $W(s)$, $s < t_1 < t_2$

N.B. Gaussian white noise can be considered an idealization of the random thermal force due to collisions with solvent molecules in the initial example of Brownian motion of a colloidal particle. in the limit correlation time $\tau_m \to 0$.

Physicist's view

$\xi(t) \equiv$ Gaussian white noise

Correlation time $\tau_c \; \sigma \longrightarrow 0. \Rightarrow$

δ-correlated noise.

$$\langle \xi(t) \rangle = 0.$$

$$\langle \xi(t_1)\, \xi(t_2) \rangle = 2D \; \delta(t_1 - t_2).$$

$$\langle \xi(t_1) \cdot \xi(t_2) \cdot \ldots \cdot \xi(t_m) \rangle$$

$$= \begin{cases} 0 & m = 2n+1 \\ \sum_{\pi} (2D)^n \, \delta(t_{i_1} - t_{i_2}) \cdots \delta(t_{i_{m-1}} - t_{i_m}) & m = 2n \end{cases}$$

- Numerical implementation of Gaussian white noise.

  - $t_j = j\,dt$.
  - $\xi_j = \xi(t_j) \sim \mathcal{N}(0, 2D/dt)$.

Example: (Euler-scheme)

$$\dot{x} = f(x,t) \cdot \xi(t)$$

$$x_j = x(t_j)$$

$$x_{j+1} = x_j + \underbrace{f(x_j, t_j) \cdot dt}_{\sim dt} + \underbrace{\xi_j\, dt}_{\sim \sqrt{dt}} + \underbrace{O(dt^{3/2})}_{\text{error}}$$

- Diffusion approximation.

  $f(t) \equiv$ Poisson process with rate r.

  $N = \int_0^{\Delta t} dt\, f(t)$.

  $\langle N \rangle = r\Delta t$, $\langle (N - r\Delta t)^2 \rangle = r\Delta t$.

  $\Rightarrow$

  $f(t) \simeq r + \xi(t)$ where

  $\xi(t) \equiv$ Gaussian white noise with

  $\langle \xi(t) \rangle = 0$.

  $\langle \xi(t)\xi(t') \rangle = r \cdot \delta(t-t')$.

N.B. Many stochastic processes can be mapped approx. on Gaussian white noise

Poisson process

$$\xi(t) = \sum_{j=-\infty}^{\infty} \delta(t - t_j)$$

$\longrightarrow t$

time: continuous

n-stage Bernoulli trial

$p = r \cdot dt$

$\longrightarrow t$

dt

time: discrete

$\Downarrow$ number of events in $[0,T]$ $\Downarrow$

Poisson distribution

$$k = \int_0^T dt \; \xi(t)$$

Binomial distribution

k: "yes"

$\Downarrow$ if $\langle k \rangle$ large

Normal approximation

$\approx N(rT, rT)$

$\approx N(np, npq)$.

Master equation (M-equation)

- Time continuous, continuous state variables

$$\frac{\partial}{\partial t} P(x,t) = \int dx_1 \; \underbrace{W(x|x_1)P(x_1,t)}_{\text{gain}} - \underbrace{W(x_1,x)P(x,t)}_{\text{loss}}$$

$W(x_2|x_1)$: transition rate density $x_1 \longrightarrow x_2$

i.e. $W(x_2|x_1)\,\Delta x_1\,\Delta x_2\,\Delta t$ = probability of transition from $[x_1, x_1+\Delta x_1]$ to $[x_2, x_2+\Delta x_2]$ in time interval Δt.

later: Fokker Planck eqn. as important special case.

- Time continuous, discrete state variables

$$\frac{d}{dt} p_n(t) = \sum_m W_{nm}\, p_m(t) - W_{mn}\, p_n(t)$$

W_{nm} = transition rate $m \to n$

Example I: Fermi's golden rule.

$H = H_0 + \varepsilon H_1 \Theta(t)$ = perturbed Hamiltonian

$|m_0\rangle$ = eigenstate of unperturbed system (H_0) = initial state

$$W_{n m_0} = \frac{2\pi}{\hbar} \varepsilon^2 |\langle n | H_1 | m_0\rangle|^2 \rho' + O(\varepsilon^3)$$

for moderate t.

van Kampen

Example II (Radioactive decay).

$n \equiv$ number of particles left.

$$T_{\Delta t}(n|m) = \begin{cases} 0 & n > m \\ m\gamma\Delta t & n = m-1 \\ O(\Delta t^2) & n < m-1. \end{cases}$$

$$\frac{d}{dt} p_n = \gamma(n+1)\, p_{n+1} - \gamma n\, p_n$$

$$N = \langle n \rangle = \sum n\, p_n.$$

$$\frac{d}{dt} N = -\gamma N \Rightarrow N = N_0 \exp -\gamma t$$