Computational statistics, lecture math 2

Frank Miller, Department of Computer and Information Science,
Linköping University

February 3, 2025

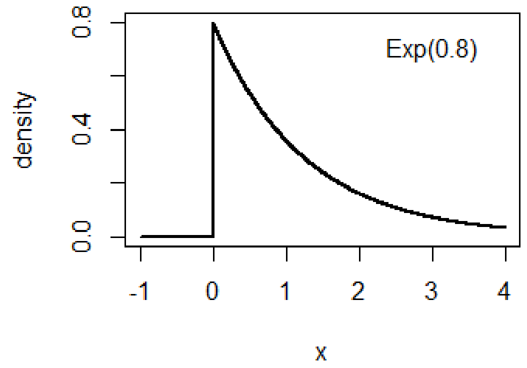
Topics today

- Probability distributions (reading: GH or Gentle, chapter 1.3) and integration
- Concavity, convexity and implication for optimization
- Markov chains (reading: GH, chapter 1.7)

Exponential distribution

- The density of the exponential distribution $Exp(\lambda)$ is

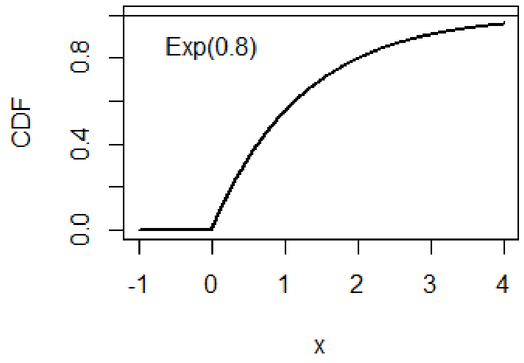
$$f(x) = \lambda \exp(-\lambda x)\, \mathbf{1}\{x \geq 0\} = \begin{cases} \lambda \exp(-\lambda x), & \text{if } x \geq 0, \\ 0, & \text{if } x < 0 \end{cases}$$




```
dexp(x, rate=0.8)
```

- Compute the cumulative distribution function (CDF):

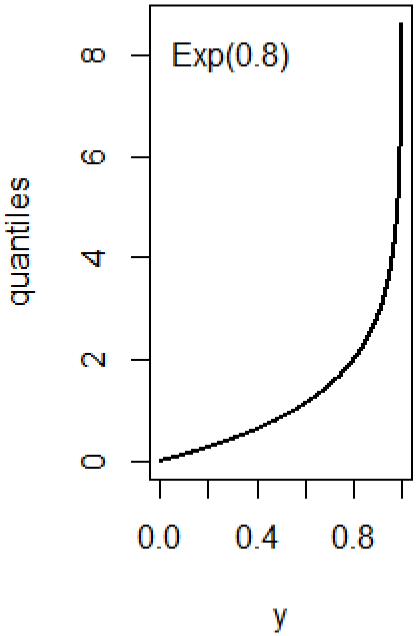
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 1 - \exp(-\lambda x), & \text{if } x \geq 0, \\ 0, & \text{if } x < 0 \end{cases}$$




```
pexp(x, rate=0.8)
```

- The inverse cumulative distribution function is $F^{-1}(y) = -\log(1-y)/\lambda$

since $y = 1 - \exp(-\lambda x) \Leftrightarrow \exp(-\lambda x) = 1 - y \;\Leftrightarrow$
$-\lambda x = \log(1-y) \Rightarrow x = -\log(1-y)/\lambda$




```
qexp(x, rate=0.8)
```

Expected value and variance of a distribution

- The probability for $X \in [a, b]$ can be written as integral according to

$$P(X \in [a,b]) = \int_a^b f(x)dx = F(b) - F(a)$$

- The expected value of a distribution with density $f(x)$ is:

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$$

- The variance is

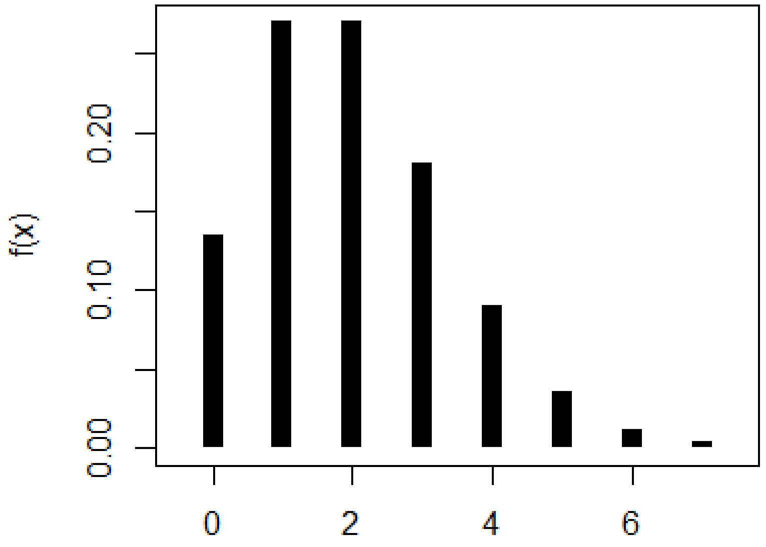
$$Var(X) = \int_{-\infty}^{\infty} (x - EX)^2\, f(x)dx$$

- What is $E(X)$ and $Var(X)$ for the exponential distribution?

Integration

- The expected value of the exponential distribution is:
  $\mathrm{E}(X) = \int_{-\infty}^{\infty} x\, f(x) dx = \int_{-\infty}^{\infty} x\lambda \exp(-\lambda x)\, \mathbf{1}\{x \geq 0\} dx =$

Poisson distribution



- The density of the Poisson distribution $Po(\lambda)$ is

$$f(x) = \frac{\lambda^x}{x!}\exp(-\lambda) \text{ for } x = 0, 1, 2, \ldots$$

`dpois(x, lambda=2)`

- The cumulative distribution function is:

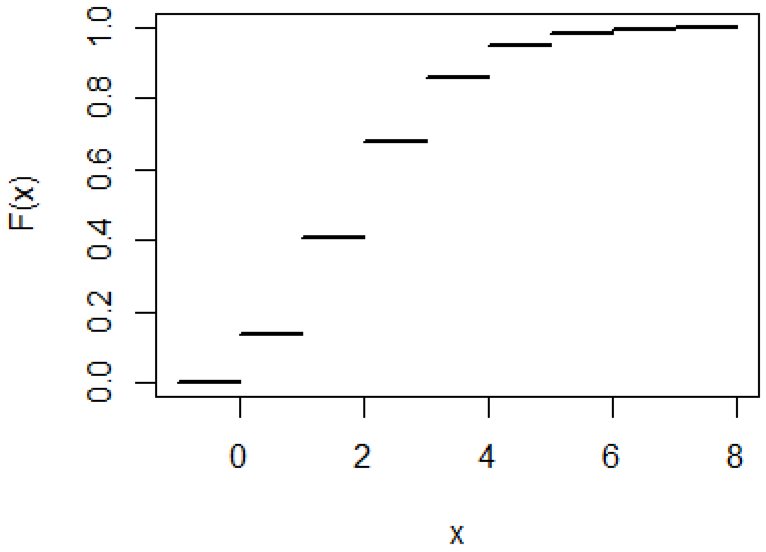
$$F(x) = P(X \leq x) = \sum_{j=0}^{x} f(j) = \exp(-\lambda)\sum_{j=0}^{x}\frac{\lambda^j}{j!}$$

`ppois(x, lambda=2)`

- The inverse CDF (quantile function) can be defined as

$$F^{-1}(y) = \min\{x: F(x) \geq y\}$$

`qpois(x, lambda=2)`



Exponential family (GH or Gentle, chapter 1.3)

- Exponential distribution: $f(x) = \lambda \exp(-\lambda x)\,\mathbf{1}\{x \geq 0\}$

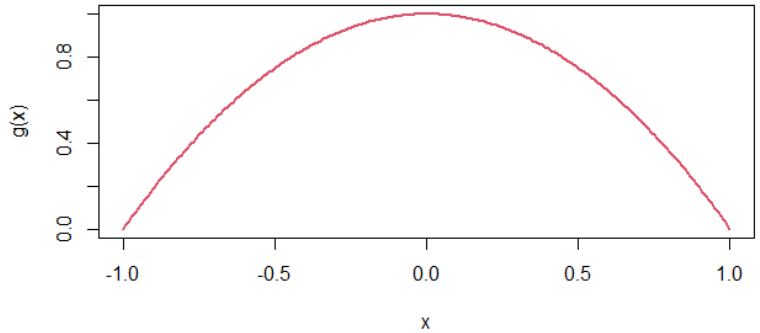
- Poisson distribution $f(x) = \frac{\lambda^x}{x!}\exp(-\lambda)$

- Exponential, Poisson, normal, … distributions belong to the exponential family of distributions
- Density of exponential family distribution:

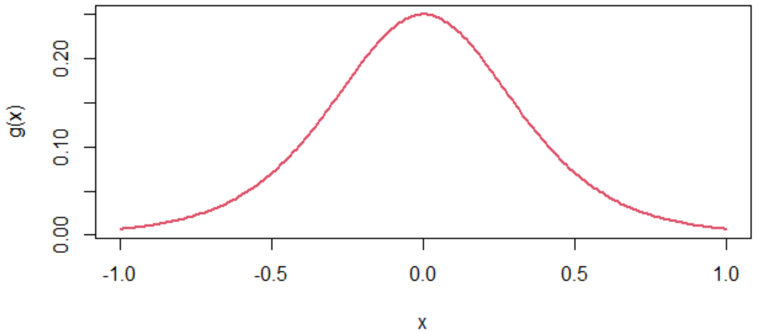
$$f(x) = c_1(x)c_2(\gamma)\exp\left(\sum_{i=1}^{k} y_i(x)\theta_i(\gamma)\right)$$

Convexity, concavity and implication for optimization

- Function g concave, if $g((\boldsymbol{x}+\boldsymbol{y})/2) \geq (g(\boldsymbol{x})+g(\boldsymbol{y}))/2$ for all $\boldsymbol{x}, \boldsymbol{y}$



concave



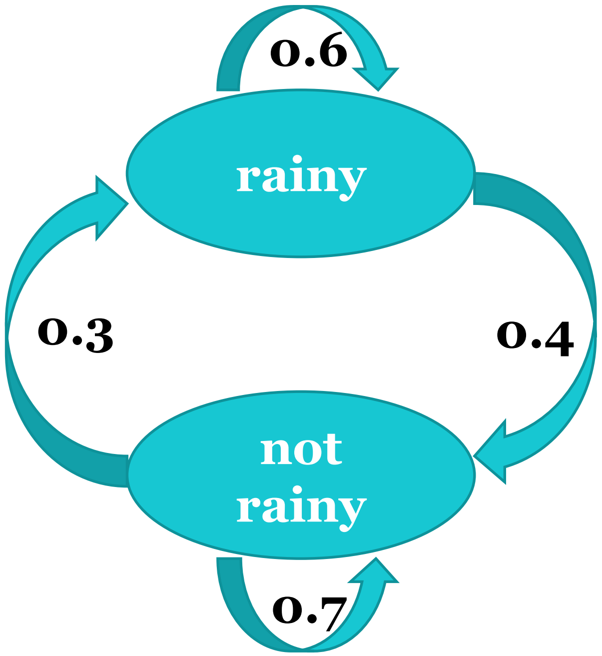
non-concave

- If g is concave, a local maximum is a global maximum
- Log likelihood for exponential families is concave
- Log likelihoods can be non-concave (e.g., Cauchy-distribution)
- Deep learning optimization problems are often non-concave / non-convex and have multiple local extrema

Markov chains (reading: GH chapter 1.7)

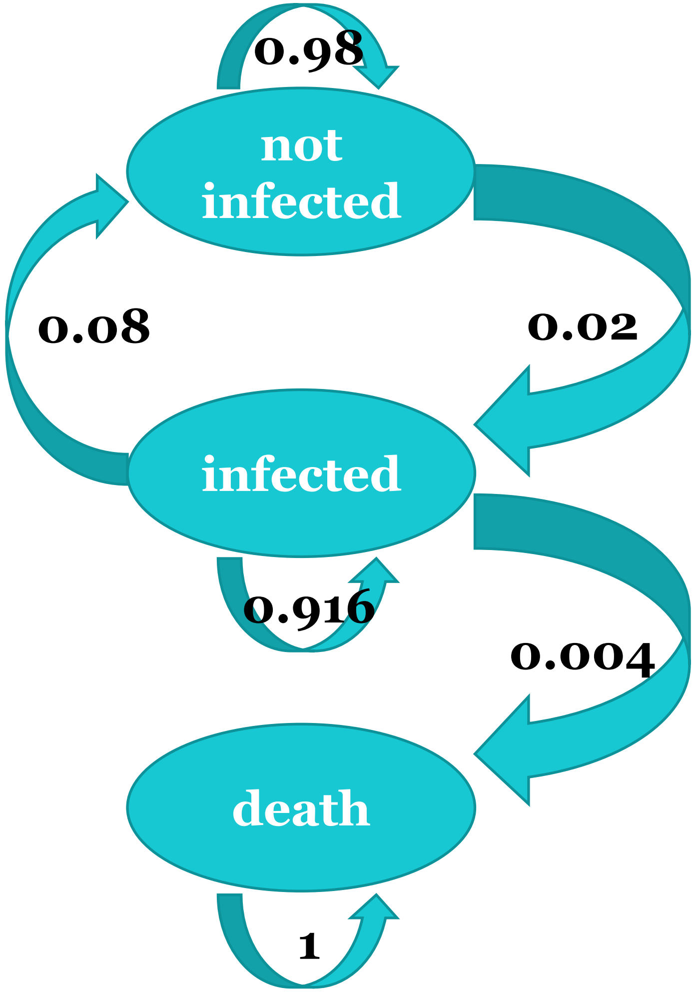
- Consider a sequence of random variables $\left(X^{(t)}\right), t = 0, 1, \dots$
- Each $X^{(t)}$ can have values in so-called *state space* S (can be discrete; we will later use a continuous one- or multidimensional S)
- A general random sequence (process) is described by specification of $P(X^{(t)}|X^{(t-1)}, \dots, X^{(0)})$ for all t

- A **Markov chain** is a specific process with *Markov property*
$$\boldsymbol{P}\left(\boldsymbol{X}^{(t)}\middle|\boldsymbol{X}^{(t-1)}, \dots, \boldsymbol{X}^{(0)}\right) = \boldsymbol{P}(\boldsymbol{X}^{(t)}|\boldsymbol{X}^{(t-1)})$$
- The state of $X^{(t)}$ depends only on the state before and not earlier history ($X^{(t)}$ is memoryless)
- $P(X^{(t)}|X^{(t-1)})$ is called *transition probability*



state space
S={rainy, not rainy}

numbers are
transition probabilities

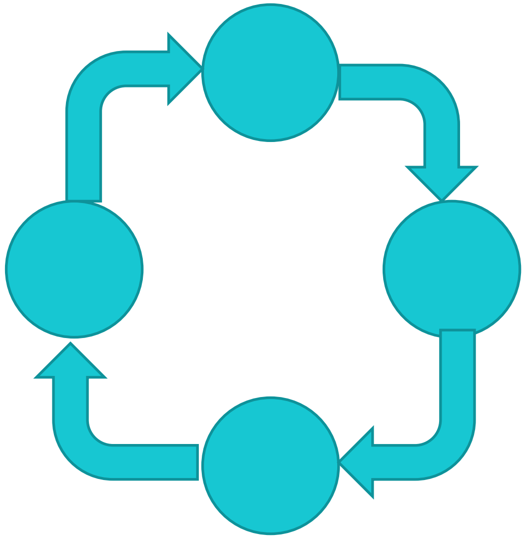


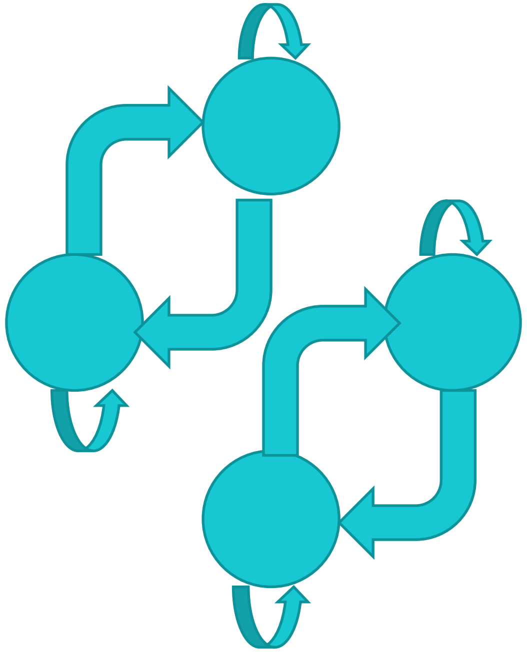
S={not infected, infected, death}

Markov chains

- If S is discrete, we can write $p_{ij}^{(t)} = P(X^{(t)} = j | X^{(t-1)} = i)$
- If S is continuous, $P(X^{(t)} | X^{(t-1)})$ is represented by a cumulative distribution function or density

- A Markov chain is *(time-)homogenous* if
  distribution of $(X^{(t)} | X^{(t-1)})$ is equal for all t
- A Markov chain is *irreducible* if every state in S can be reached from any state
- A state has *period* k if multiples of k steps are necessary to return to it
- A Markov chain is *aperiodic* if each state has period 1





- A Markov chain is *irreducible* if every state in S can be reached from any state
- A state has *period* k if multiples of k steps are necessary to return to it
- A Markov chain is *aperiodic* if each state has period 1