Simulation

Alberto Ceselli
MSc in Computer Science
Univ. of Milan

Part 1 - Modeling with probabilities

Probability Recap

Recall:

- Sample Space S (e.g. outcome of a horse race)
- Event (e.g. arrival $(1,5,2,4,3)$).
- Union and Intersection of Events
- Complement of an Event
- Mutually exclusive events

Axioms of Probability

For every Event A, the *probability* of A ($P(A)$) is a number s.t.

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- For a sequence of mutually exclusive events $A_1 \ldots A_N$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

- E.g. $P(\bar{A}) = 1 - P(A)$

Conditional Probability

Example: flipping a coin twice

- what is the probability of (H,H)?
- what is the probability of (H,H) if you know (*conditioning*) the first flip was H?
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
- $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$
- Examples 2a and 2b: insurance company (page 8).

A modeler's view of Variables

$$y = f(x)$$

How many "variables" do you see?

A modeler's view of Variables

$$y = f(x)$$

How many "variables" do you see? Intuitively, for a *modeler*, x is *input data*, while y is **an abstraction of the process modeled by** $f()$

Random Variables

Example: you have an experiment, whose output is X

- i.e. X is a *variable*, assuming to contain the result of your experiment
- if the experiment involves some stochastic behavior, the content of X will always be uncertain
- therefore, X is called *random* variable
- they can be *discrete* or *continuous*

Variables and Functions

- C.D.F. is $F(x) = P[X \leq x]$
- if X is discrete, its *probability mass function* is $p(x) = P[X = x]$
- $\sum_{i \in I} p(x_i) = 1$
- X is continuous, if a (non negative) *probability density function* $f(x)$ exists, such that, for $C \subseteq \mathbb{R}$ is $P[X \in C] = \int_C f(x)dx$

Variables and Functions

That is,

- $F(a) = P[X \in (-\infty, a)] = \int_{-\infty}^{a} f(x)dx$
- or equivalently $\dfrac{dF(x)}{dx} = f(x)$

Joint Probabilities

When we have more than one random variable

- $F(x, y) = P[X \leq x \wedge Y \leq y]$
- $p(x, y) = P[X = x \wedge Y = y]$
- X and Y are *independent* if, for any pair of sets $C \subseteq \mathbb{R}$, $D \subseteq \mathbb{R}$, it holds
$$P[X \in C, Y \in D] = P[X \in C] \cdot P[Y \in D]$$
- i.e. for any choice of C and D, the events $X \in C$ and $Y \in D$ are independent

Expectation

If X is a discrete random variable taking values $x_1 \ldots x_n$

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

If X is a continuous random variable with p.d.f. f

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

What if we need to compute $E[g(X)]$, being $g()$ a generic function?

Expectation

If X is a discrete random variable taking values $x_1 \ldots x_n$

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

If X is a continuous random variable with p.d.f. f

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

What if we need to compute $E[g(X)]$, being $g()$ a generic function?
The case $aX + b$ (proof on the whiteboard)

Expectation

If X is a discrete random variable taking values $x_1 \ldots x_n$

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

If X is a continuous random variable with p.d.f. f

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

What if we need to compute $E[g(X)]$, being $g()$ a generic function?
The case $aX + b$ (proof on the whiteboard)

The general case: $E[g(X)] = \sum_{i=1}^{n} g(x_i) \cdot p(x)$

Expectation

If X is a discrete random variable taking values $x_1 \ldots x_n$

$$E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$$

If X is a continuous random variable with p.d.f. f

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

What if we need to compute $E[g(X)]$, being $g()$ a generic function?
The case $aX + b$ (proof on the whiteboard)

The general case: $E[g(X)] = \sum_{i=1}^{n} g(x_i) \cdot p(x)$

Expectation is a *linear* operator (proof on the whiteboard)

Variance

If X is a discrete random variable with expectation μ

- $Var[X] = E[(X-\mu)^2] = E[X^2] - \mu^2$
- $Var[aX+b] = a^2 Var[X]$
- Var is **not** a linear operator
- $Cov[X,Y] = E[(X-\mu_x)\cdot(Y-\mu_y)] = E[X\cdot Y] - E[X]\cdot E[Y]$
- What's the meaning of covariance?
- $Var[X+Y] = Var[X] + Var[Y] - 2Cov[X\cdot Y]$
- So, when X and Y are independent ...
- $Corr[X,Y] = Cov[X,Y]/\sqrt{Var[X]\cdot Var[Y]}$

It is a matter of size

Markov's inequality (proof on the whiteboard): if X takes only nonnegative values, for any $a > 0$

$$P[X \geq a] \leq E[X]/a$$

Chebishev's inequality (proof on the whiteboard): if X takes only nonnegative values, has expectation μ and variance σ^2, then for any $k > 0$

$$P[|X - \mu| \geq k \cdot \sigma] \leq \frac{1}{k^2}$$

Weak law of large numbers (proof on the whiteboard): let $X_1 \dots X_n$ be a sequence of i.i.d. random variables having expectation μ. Then, for any $\epsilon > 0$

$$P[|\frac{\sum_{i=1}^{n} X_i}{n} - \mu| > \epsilon]] \to 0 \text{ as } n \to \infty$$

Strong law of large numbers: $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i}{n} = \mu$ with probability 1

Binomial Random Variables

Application: Alice and Bob play a dice game. It consists in rolling a single dice **exactly** 10 times: Alice wins if she gets 5 times the value 1. What is her probability of winning?

- Number of events known in advance (n)
- They are *independent* one another
- They have *equal probability of success* (p)

$\rightarrow$ binomial random var X = number of successes.

$$P[X = i] = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$$

Poisson Random Variables

Application: Alice and Bob own a computers shop. Alice: ”Wow, today there’s nobody around: I usually see about 20 customers every morning”. Bob: ”Don’t worry. I bet tomorrow morning we’ll have 40 of them!”. What’s the probability for Bob to win his bet? What’s the probability for the shop to remain empty during all the morning?

- Very large number of events.
- They are *independent* one another
- The *expected number of positive ones* in known in advance (λ)

$\rightarrow$ poisson random var X = number of successes.

$$P[X = i] = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

Poisson Random Variables

Application: Alice and Bob own a computers shop. Alice: ”Wow, today there’s nobody around: I usually see about 20 customers every morning”. Bob: ”Don’t worry. I bet tomorrow morning we’ll have 40 of them!”. What’s the probability for Bob to win his bet? What’s the probability for the shop to remain empty during all the morning?

- Very large number of events.
- They are *independent* one another
- The *expected number of positive ones* in known in advance (λ)

$\rightarrow$ poisson random var X = number of successes.

$$P[X = i] = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$

Geometric Random Variables

Application: Alice and Bob are still bored. Alice: ”Hey, that dice game was cool! Let’s play another one. You roll the dice until you get 6!”. Bob: ”Great! I bet I can win in exactly 3 rolls!”. What is his probability of winning?

- Number of events is irrelevant
- They are *independent* one another
- They have *equal probability of success* (p)
- Only first success counts.

$\rightarrow$ geometric random var X = first trial with success.

$$P[X = i] = p \cdot (1 - p)^{i-1}$$

Negative Binomial Random Variables

Application: Alice: "That was too easy! Now, you roll the dice until you get 6 *twice*!". Bob: "Great! I bet I can win in at most 5 rolls!". What is his probability of winning?

- Number of events is irrelevant
- They are *independent* one another
- They have *equal probability of success* (p)
- There is a *target number of successes* (r)

$\rightarrow$ negative binomial random var X = number of trials to get r successes: for $i \geq r$

$$P[X = i] = \binom{i-1}{r-1} p^r \cdot (1-p)^{i-r}$$

Hypergeometric Random Variables

Application: Alice: ”Hey Bob: I have found a beautiful deck of 52 cards below the counter! Now I shuffle, and you pick 4 of them: I bet you can't get 2 red and 2 blue!”. What is her probability of winning?

- A population is given, where N individuals hold a feature and M do not.
- Size of a sample is given (n)

$\rightarrow$ hypergeometric random var X = number of individuals in the sample holding the feature

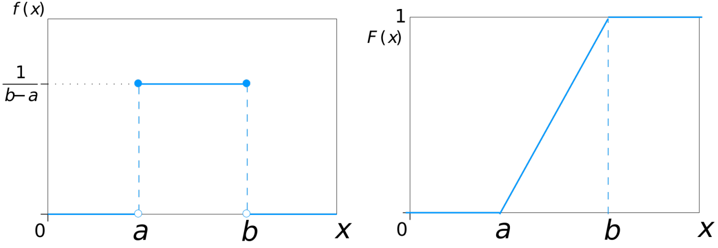
$$P[X = i] = \frac{\binom{N}{i} \cdot \binom{M}{n-i}}{\binom{N+M}{n}}$$

Exercises

Number 15, (18, 23), 36.

Uniformly Distributed Random Variables

Parameters: range $[a, b]$. Visually, [1]



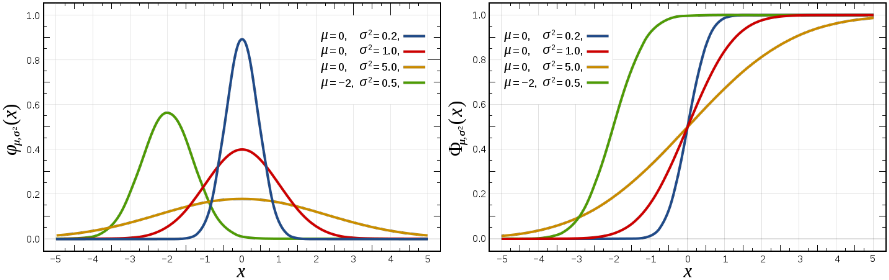
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \qquad F(x) = \frac{x-a}{b-a}$$

[1] IkamusumeFan - Own work. This drawing was created with LibreOffice Draw, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=27378784

Normal Random Variables

Parameters:

expectation μ, variance σ^2. Visually,[2]



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad F(x) = \Phi(\frac{x-\mu}{\sigma^2}) \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-z^2/2}\, dz$$

[2]By Inductiveload - self-made, Mathematica, Inkscape, Public Domain, https://commons.wikimedia.org/w/index.php?curid=3817960

Normal Random Variables

- $\Phi(x)$ cannot be expressed with only additions, subtractions, multiplications, and root extractions (i.e. need for numerical evaluation / approximation)
- if X is normally distributed with param. μ and σ^2, $Z = aX + b$ is normally distributed with param. $a\mu + b$ and $a^2\sigma^2$ (e.g. if $a = 1/\sigma^2$ and $b = -\mu/\sigma^2$ ...)
- **Central Limit Theorem.** Let $X_1 \dots X_n$ be a sequence of n i.i.d. random variables having finite exp. value μ and variance σ^2

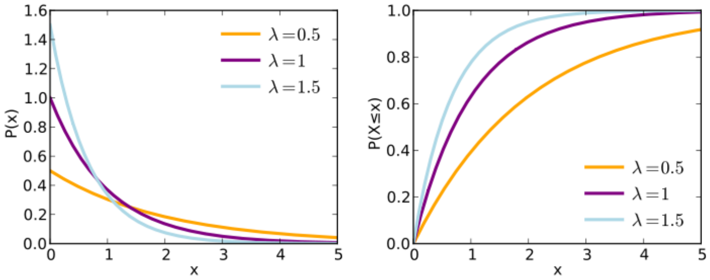
$$\lim_{n\to\infty} P[\frac{\sum_{i=1}^{n} X_i - n \cdot \mu}{\sqrt{n\sigma^2}} < x] = \Phi(x)$$

i.e. the sum of a large number of independent random variables is a normally distributed random variable *(independently on the distribution of the starting ones)*

Exponential Random Variables

Parameters:

rate λ (expectation $1/\lambda$). Visually,[3]



For $0 < x < \infty$

$$f(x) = \lambda \cdot e^{-\lambda \cdot x} \qquad F(x) = 1 - e^{-\lambda \cdot x}$$

[3]By Skbkekas - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=9508326

Exponential Random Variables

- The *only ones* having memoryless property (e.g. X is the lifetime of an item):
$$P[X > s + t | X > s] = P[X > t]$$
- They remain exponential when multiplied by positive constants: if X is exponential with parameter λ $Y = cX$ is exponential with parameter λ / c (proof on the whiteboard)
- If $X_1 \dots X_n$ are independent exponential random variables, $Y = \min_i X_i$ is exponential with rate $\sum_i \lambda_i$ (proof on the whiteboard)
- The probability that X_j is the smallest is $\lambda_j / \sum_i \lambda_i$ (proof on the whiteboard)

Poisson Processes

A **Poisson Process** having **rate** λ is:

- *Events* are occurring at random time points
- $N(t)$ is the number of events in the interval $[0, t]$
- $N(0) = 0$ (process begins at time 0)
- Number of events in disjoint time intervals are independent (independent increment assumption)
- The PDF of the number of events in a given interval depend only on its length, not on its position (stationary increment assumption)
- $\lim_{h \to 0} \frac{P[N(h) = 1]}{h} = \lambda$ (in small intervals, the probability of an event to occur is approximately $h\lambda$)
- $\lim_{h \to 0} \frac{P[N(h) \geq 2]}{h} = 0$ (unlikely that two or more events occur in small intervals)

Poisson Processes

- **Claim:** $N(t)$ is a Poisson Random Variable with expected value $t\lambda$ (proof on the whiteboard)
- **Claim:** The interarrival times are i.i.d. exponential random variables with parameter λ
- **Definition:** A *Gamma Random Variable* with parameters n, λ is a (continuous) random variable having the following PDF:

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}$$

- **Claim:** The sum of n independent exponential random variables, each having parameter λ, is a gamma random variable with parameters n, λ
- **Claim:** The time of the n-th event of a Poisson process having rate λ is a gamma random variable with parameters n, λ

Nonhomogeneous Poisson Processes

A **Poisson Process** having **rate** λ is:

- *Events* are occurring at random time points
- $N(t)$ is the number of events in the interval $[0, t]$
- $N(0) = 0$ (process begins at time 0)
- Number of events in disjoint time intervals are independent (independent increment assumption)
- PDF of the number of events in a given interval depend only on its length, not on its position (stationary increment assumption)
- $\lim_{h \to 0} \frac{P[N(h) = 1]}{h} = \lambda$ (in small intervals, the probability of an event to occur is approximately $h\lambda$)
- $\lim_{h \to 0} \frac{P[N(h) \geq 2]}{h} = 0$ (unlikely that two or more events occur in small intervals)

Nonhomogeneous Poisson Processes

A **Poisson Process** having **rate** λ is:

- *Events* are occurring at random time points
- $N(t)$ is the number of events in the interval $[0, t]$
- $N(0) = 0$ (process begins at time 0)
- Number of events in disjoint time intervals are independent (independent increment assumption)
- PDF of the number of events in a given interval depend only on its length, not on its position (stationary increment assumption)
- $\lim_{h \to 0} \frac{P[N(h) = 1]}{h} = \lambda$ (in small intervals, the probability of an event to occur is approximately $h\lambda$)
- $\lim_{h \to 0} \frac{P[N(h) \geq 2]}{h} = 0$ (unlikely that two or more events occur in small intervals)

Nonhomogeneous Poisson Processes

A Nonhomogeneous **Poisson Process** having **rate** $\lambda(t)$ is:

- *Events* are occurring at random time points
- $N(t)$ is the number of events in the interval $[0, t]$
- $N(0) = 0$ (process begins at time 0)
- Number of events in disjoint time intervals are independent (independent increment assumption)
- $\lambda(t)$ is the *intensity* at time t (i.e. how likely an event will occur around time t)
- $\lim_{h \to 0} \frac{P[N(h) = 1]}{h} = \lambda(t)$ (in small intervals, the probability of an event to occur is approximately $h\lambda$)
- $\lim_{h \to 0} \frac{P[N(h) \geq 2]}{h} = 0$ (unlikely that two or more events occur in small intervals)

Nonhomogeneous Poisson Processes

Interpretation: let $\bar{p}(t)$ be the *probability* that an event occurring at time t in a Poisson process with parameter λ is *discarded*, and $p(t) = 1 - \bar{p}(t)$; then the process involving the non-discarded events is a nonhomogeneous Poisson Process with intensity $\lambda(t) = \lambda \cdot p(t)$

- Let
$$m(t) = \int_0^t \lambda(s)ds$$
- $Y = N(t+s) - N(t)$ is a Poisson Random Variable with expected value $m(t+s) - m(t)$