

Applied Probability Formulary

CA' FOSCARI UNIVERSITY OF VENICE
Department of Environmental Sciences, Informatics and Statistics



[CM0546] APPLIED PROBABILITY FOR COMPUTER SCIENCE
Academic Year 2022 - 2023

Student Zuliani Riccardo 875532

Contents

1	Function Properties	1
1.1	Mean	1
1.2	Variance	1
1.3	Covariance	1
2	Discrete RVs	2
2.1	Bernoulli Distribution	2
2.2	Binomial Distribution	2
2.3	Geometric Distribution	3
2.4	HyperGeometric Distribution	3
2.5	Multinomial Distribution	4
2.6	Negative Binomial Distribution	5
2.7	Poisson Distribution	5
2.7.1	Poisson approximation of Binomial distribution	6
2.7.2	Additivity	6
2.7.3	Relation between Poisson and Multinomial Distribution	6
3	Continuous RVs	7
3.1	Uniform Distribution	7
3.2	Exponential Distribution	7
3.2.1	Times between rare events are Exponential	8
3.2.2	Memory Less Propriety	8
3.2.3	Minimization	8
3.2.4	Maximization	8
3.3	Gamma Distribution	8
3.4	Normal Distribution	9
3.4.1	Standard Normal Distribution	9
3.5	Central Limit Theorem	10
3.5.1	Normal Approximation to Binomial Distribution	11
3.5.2	Continuity Correction	11
4	Approximation	12
5	Stochastic Processes	13
5.1	Proprieties	13
5.1.1	Mean Function	13
5.1.2	Variance Function	13
5.1.3	Standard Deviation Function	14
5.1.4	Auto-Covariance Function	14

5.1.5	Auto-Correlation Function	14
5.2	Stationary and wide-sense stationary processes	14
5.2.1	Strongly / Strict-Sense Stationary	14
5.2.2	Weakly / Wide-Sense Stationary	15
5.3	Markov Processes	16
5.3.1	Markov Chain	16

Chapter 1

Function Properties

1.1 Mean

- $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
- $\mathbb{E}[cX] = c\mathbb{E}[X]$
- **For independent X and Y** $\mathbb{E}[XY] = \mathbb{E}[X] \times \mathbb{E}[Y]$
- X_1, X_2, \dots, X_n RVs, then $\mathbb{E}[X_1 + X_2 + \dots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] = \sum_{i=1}^n \mathbb{E}[X_i]$

1.2 Variance

- $\text{VAR}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
- $\text{VAR}[X + c] = \text{VAR}[X]$ $\text{VAR}[cX] = c^2\text{VAR}[X]$
- $\text{VAR}[aX + bY + c] = a^2\text{VAR}[X] + b^2\text{VAR}[Y] + 2ab\text{COV}[X, Y]$
- **For independent X and Y** $\text{VAR}[X + Y] = \text{VAR}[X] + \text{VAR}[Y]$

1.3 Covariance

- $\text{VAR}[X] = \text{COV}[X, X]$
- $\text{COV}[cX, Y] = c\text{COV}[X, Y]$ $\text{COV}[X, cY] = c\text{COV}[X, Y]$
- $\text{COV}[X + Y, Z] = \text{COV}[X, Z] + \text{COV}[Y, Z]$
- $\text{COV}[X, Y + Z] = \text{COV}[X, Y] + \text{COV}[X, Z]$
- $\text{COV}[X, Y] = \text{COV}[Y, X]$
- $\text{COV}[X, c] = 0$
- **For independent X and Y** $\text{COV}[X, Y] = 0$
- $\text{COV}[X + Y, Z + W] = \text{COV}[X, Z] + \text{COV}[X, W] + \text{COV}[Y, Z] + \text{COV}[Y, W]$

Chapter 2

Discrete RVs

2.1 Bernoulli Distribution

Used whenever we have a 0 / 1 outcome, thus when we could have only two possible result in our experiment.

p	probability of success
$P[X]$	$p^x(1-p)^{1-x} = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \end{cases}$
$E[X]$	p
$VAR[X]$	$p(1-p)$

2.2 Binomial Distribution

Used whenever we consider a sequence of independent Bernoulli trials and count the number of success in it.

n	number of trials
p	probability of success
$P[x]$	$\binom{n}{x} p^x (1-p)^{n-x}$
$F[x]$	$\sum_{i=1}^n \binom{n}{i} p^i (1-p)^{n-i}$
$E[X]$	np
$VAR[X]$	$np(1-p)$
$P[X = x]$	<code>dbinom(<i>#success</i> , <i>size</i> , <i>prob_success</i>)</code>
$P[X \leq x]$	<code>pbinom(<i>#success</i> , <i>size</i> , <i>prob_success</i>)</code>

2.3 Geometric Distribution

Consider a sequence of independent Bernoulli trials, each trial results in a "success" or a "failure". The number of Bernoulli trials needed to get the first success has Geometric Distribution.

p	probability of success
$P[x]$	$(1-p)^{x-1}p, x = 1, 2, \dots$
$F[x]$	$p \sum_{i=0}^x (1-p)^i$
$\mathbf{E}[X]$	$\frac{1}{p}$
$\text{VAR}[X]$	$\frac{1-p}{p^2}$
$P[X = x]$	dgeom (#failures , prob_success)
$P[X \leq x]$	pgeom (#failures , prob_success)

2.4 HyperGeometric Distribution

Describes the probability of k successes (random draws for which the object drawn has a specified feature) in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or a failure.

N	Is the population size
K	Is the number of success states in the population
n	Is the number of draws
k	Is the number of observed successes
$P[x]$	$\frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$
$\mathbf{E}[X]$	$n \frac{K}{N}$
$P[X = x]$	dhyper (#succ , #succ_samp , #pop_dim , #samp_dim)
$P[X \leq x]$	phyper (#succ , #succ_samp , #pop_dim , #samp_dim)

2.5 Multinomial Distribution

Whereas the binomial distribution describes the number of successes in a Bernoulli process, for which each single test can provide only two results, the multinomial distribution describes the more general case in which each test can provide a finite number of results, each with the own probability.

$$P[X_1, X_2, \dots, X_k] = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k} \quad \text{where} \quad \sum_{i=1}^n p_i = 1$$

2.6 Negative Binomial Distribution

In a sequence of independent Bernoulli trials, the number of trials needed to obtain k successes has Negative Binomial distribution.

In other words it counts the number of failures before obtaining a target number of successes (k).

k	number of success
p	probability of success
$P[x]$	$\binom{x-1}{k-1} (1-p)^{x-k} p^k \quad x = k, k+1, \dots$
$\mathbf{E}[X]$	$\frac{k}{p}$
$\text{VAR}[X]$	$\frac{k(1-p)}{p^2}$
$P[X = x]$	<code>dnbinom(#failures, #successes, prob_success)</code> <code>dnbinom(#trial-#successes, #successes, prob_success)</code>
$P[X \leq x]$	<code>pnbinom(#failures, #successes, prob_success)</code> <code>pnbinom(#trial-#successes, #successes, prob_success)</code>

2.7 Poisson Distribution

Poisson distribution is related to **rare events**. It means that two events are extremely unlikely to occur simultaneously or within a very short period of time.

The number of rare events occurring within a fixed period of time has Poisson Distribution.

λ	frequency, average number of events
p	probability of success
$P[x]$	$e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots$
$\mathbf{E}[X]$	λ
$\text{VAR}[X]$	λ
$P[X = x]$	<code>dpois(x, lambda)</code>
$P[X \leq x]$	<code>ppois(x, lambda)</code>

2.7.1 Poisson approximation of Binomial distribution

Poisson distribution can be effectively used to approximate Binomial probabilities when the *number of trials* n is **large** and the *probability of success* p is **small**

$$n \geq 30 \quad p \leq 0.05$$

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda)$$

$$\text{where } n \geq 30 \quad p \leq 0.05 \quad np = \lambda$$

2.7.2 Additivity

If

$$X \sim \text{Pois}(\lambda) \quad \text{and} \quad Y \sim \text{Pois}(\mu)$$

and they are **independent**, then we can say that:

$$W = X + Y \sim \text{Pois}(\lambda + \mu)$$

2.7.3 Relation between Poisson and Multinomial Distribution

Let

$$S_n = X_1 + X_2 + \dots + X_n \quad \text{with} \quad X_i \stackrel{iid}{\sim} \text{Pois}(\lambda_i)$$

given

$$(X_1, X_2, \dots, X_n) | S_n \sim \text{Mult}\left(\frac{\lambda_1}{\lambda}, \frac{\lambda_2}{\lambda}, \dots, \frac{\lambda_n}{\lambda}\right) \quad \text{where} \quad \lambda = \sum_{i=1}^n \lambda_i$$

Chapter 3

Continuous RVs

3.1 Uniform Distribution

The distribution describes an experiment where there is an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters, a and b , which are the minimum and maximum values.

(a, b)	range of values
$f(x)$	$\frac{1}{b-a} \quad a < x < b$
$F_x(x)$	$\frac{x-a}{b-a} \quad a < x < b$
$\mathbf{E}[X]$	$\frac{a+b}{2}$
$\text{VAR}[X]$	$\frac{(b-a)^2}{12}$
$P[X = x]$	dunif (x, min, max)
$P[X \leq x]$	punif (x, min, max)

3.2 Exponential Distribution

Exponential distribution used to model **time**. In a sequence of rare events, when the number of events is Poisson, the time between events is Exponential

λ	frequency parameter, the number of events per time unit
$f(x)$	$\lambda e^{-\lambda x} \quad x > 0$
$F_x(x)$	$1 - e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{1}{\lambda}$
$\text{VAR}[X]$	$\frac{1}{\lambda^2}$
$P[X = x]$	dexp (x, rate ($\wedge -1$))
$P[X \leq x]$	pexp (x, rate ($\wedge -1$))

If in the exercise it is not explicit the measure unit at $^{-1}$, we have to insert in the **rate** parameter: $1/\text{rate} = \text{rate}^{-1}$

3.2.1 Times between rare events are Exponential

Event: "the time T until the next event is greater than t " can be rephrased as: "zero events occur by the time t ".

$$P_X(0) = e^{-\lambda t} \frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

Then the cdf of T is:

$$F_T[t] = 1 - P[T > t] = 1 - P[T = t] = 1 - e^{-\lambda t}$$

3.2.2 Memory Less Propriety

The fact of having waited for t minutes gets "forgotten", and it does not affect the future waiting time.

$$P[T > t + x | T > t] = P[T > x] \quad \forall t, x > 0$$

3.2.3 Minimization

Consider a collection of $X_j \sim \text{Exp}(\lambda_j)$ with $j = 1, \dots, n$ independent from each other we state that there exist a new random variable:

$$L_n = \min\{X_1, \dots, X_n\} \sim \text{Exp}(\lambda) \quad \lambda = \sum_{j=1}^n \lambda_j$$

It has the same propriety as a classical **Poisson Random Variable**

3.2.4 Maximization

$$\mathbb{P}[x \leq X] = \prod_{i=1}^n (1 - e^{-\lambda_i x})$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \sum_{i=1}^n \frac{1}{i}$$

3.3 Gamma Distribution

When a certain procedure consist of α *independent steps*, and each step takes **Exponential**(λ) amount of time, then the total time has **Gamma distribution** with parameters α and λ .

In a process of rare events, with **Exponential** times between any two consecutive events, the time of the α -th events has **Gamma** distribution because it consists of α *independent Exponential* times.

α	shape parameter
λ	frequency parameter
$f(x)$	$\frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad x > 0$
$\mathbf{E}[X]$	$\frac{\alpha}{\lambda}$
$\text{VAR}[X]$	$\frac{\alpha}{\lambda^2}$
$P[X = x]$	dgamma (x, alpha, rate ($\wedge -1$))
$P[X \leq x]$	pgamma (x, alpha, rate ($\wedge -1$))

If in the exercise it is not explicit the measure unit at $^{-1}$, we have to insert in the **rate** parameter: $1/\text{rate} = \text{rate}^{-1}$

3.4 Normal Distribution

Besides sums, averages, and errors, Normal distribution is often found to be a good model for physical variables like weight, height, temperature, voltage, pollution level, and for instance, household incomes or student grades.

μ	expectation, location parameter
σ	standard deviation, scale parameter
$f(x)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad -\infty < x < \infty$
$F_x[X]$	$\int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\} dz \quad -\infty < x < \infty$
$\mathbf{E}[X]$	μ
$\text{VAR}[X]$	σ^2

3.4.1 Standard Normal Distribution

Normal distribution with “standard parameters” $\mu = 0$ and $\sigma = 1$ is called **Standard Normal distribution**.

μ	expectation, location parameter
σ	standard deviation, scale parameter
Z	Standard Normal Random Variable
$\phi(x)$	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ Standard Normal pdf
$\Phi(x)$	$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ Standard Normal cdf
$P[X = x]$	dnorm ((X - mean) / sd)
$P[X \leq x]$	pnorm ((X - mean) / sd)
$\Phi^{-1}(x)$	qnorm (x)

A Standard Normal can be obtained from a non standard Normal(μ, σ) random variable X by **standardizing**, which means *subtracting* the **mean** and *dividing* by the **standard deviation**:

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

Using the transformation, any Normal Random Variable can be obtained from a **Standard Normal Random Variable** Z :

$$F_x[x] = P[X \leq x] = P\left[\frac{X' - \mu'}{\sigma'} \leq \frac{x - \mu}{\sigma}\right] = P\left[Z \leq \frac{x - \mu}{\sigma}\right] = F_z\left[\frac{x - \mu}{\sigma}\right]$$

Linear Combination of Normal RVs are Normal

$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

$$\text{and if } a_i = \frac{1}{n} \quad \forall i \quad \text{or} \quad \frac{1}{n} \sum_{i=1}^n x_i \sim N(\mu, \sigma^2/n)$$

$$\sum_{i=1}^n a_i x_i \sim N\left(\mu \sum_{i=1}^n a_i, \sigma^2 \sum_{i=1}^n a_i\right)$$

3.5 Central Limit Theorem

To be used when in the exercise it is asked to find some kind of probability given quantity of elements and relative mean and sd.

$$X_1, X_2, \dots \quad \text{independent RVs} \quad \mu = \mathbf{E}[X_i] \quad \sigma = \text{Std}[X_i]$$

$$S_n = \sum_{i=1}^n X_i = X_1 + \dots + X_n$$

As $n \rightarrow \infty$ the standardized sum is

$$Z_n = \frac{S_n - \mathbf{E}[S_n]}{\text{Std}[S_n]} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution to a **Standard Normal Random Variable**

$$F_{Z_n}(z) = P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z\right] \rightarrow \Phi(z) \quad \forall z$$

Applied when $n \geq 30$

3.5.1 Normal Approximation to Binomial Distribution

Binomial Variables represent a special case of $S_n = X_1 + \dots + X_n$, where all $X_i \sim \text{Ber}(p)$, moreover in case our n is **large** and for moderate values of p : ($0.05 \leq p \leq 0.95$) we have the following approximation

$$\text{Binomial}(n, p) \approx \text{Normal}(\mu = np, \sigma = \sqrt{np(1-p)})$$

3.5.2 Continuity Correction

It is needed when we approximate a discrete distribution (like Binomial) by a continuous distribution (Normal). Since in the discrete case $P[X = x]$ could be positive in the continuous case it is always 0. This is way we introduce this correction.

We expand the interval by 0.5 units in each direction, then use the Normal approximation.

$$P_X[x] = P[X = x] = P[x - 0.5 < X < x + 0.5]$$

Chapter 4

Approximation

Given X_1, X_2, \dots, X_n a sequence of independent random variables with $S_n = \sum_{i=1}^n X_i$ if:

- $X_i \stackrel{iid}{\sim} \text{Bernulli}(p) \approx S_n \sim \text{Binomial}(n, p)$
- $X_i \stackrel{iid}{\sim} \text{Geometric}(p) \approx S_n \sim \text{NegativeBinomial}(n, p)$
- $X_i \stackrel{iid}{\sim} \text{Exponential}(\lambda) \approx S_n \sim \text{Gamma}(\alpha = n, \lambda)$
- $X_i \stackrel{iid}{\sim} \text{Poisson}(\lambda) \approx S_n \sim \text{Poisson}(n * \lambda)$

Chapter 5

Stochastic Processes

A **Stochastic Process** is a Random Variable that also depends on **time**, thus it is a function of two arguments $X(t, w)$ where:

- $t \in \mathcal{T}$ is **time**
- $w \in \Omega$

Moreover:

- At any time t we see a **random variable** $X_t(w)$: function of random income.
- At a given w we obtain a **function of time** $X_w(t)$

We have two classification of stochastic processes and are the following one:

- **Variable Classification:**
 - $X(t, w)$ is a **discrete-state process** if X_t is a *discrete rv* $\forall t$
 - $X(t, w)$ is a **continuous-state process** if X_t is a *continuous rv* $\forall t$
- **Time Dimension Classification:**
 - $X(t, w)$ is a **discrete-state process** if the set of time \mathcal{T} is *discrete*
 - $X(t, w)$ is a **continuous-state process** if \mathcal{T} is unbounded and thus *continuous*

5.1 Proprieties

5.1.1 Mean Function

$$\mu_x(t) = \mathbb{E}[X(t)]$$

Where $\mathbb{E}[X(t)]$ is the **expected value** of the rv for the fixed *time point* t

5.1.2 Variance Function

$$\sigma^2(t) = VAR[X(t)] = \mathbb{E}[(X(t) - \mu_x(t))^2] = \mathbb{E}[X^2(t)] - [\mu_x(t)]^2$$

5.1.3 Standard Deviation Function

$$\sigma_x(t) = \sqrt{VAR[X(t)]} = \sqrt{\sigma_x^2(t)}$$

5.1.4 Auto-Covariance Function

$$\begin{aligned}\sigma_x(t, s) = C_{x,x} &= COV[X(t), X(s)] = \mathbb{E}\left[\left(X(t) - \mu_x(t)\right) \times \left(X(s) - \mu_x(s)\right)\right] \\ &= \mathbb{E}\left[\left(X(t) \times X(s)\right) - \left(\mu_x(t) \times \mu_x(s)\right)\right]\end{aligned}$$

And it has the following proprieties:

- $C_{x,x}(t, s) = C_{x,x}(s, t)$
- $\sigma_x^2(t) = VAR[X(t)] = COV[X(t), X(t)] = C_{x,x}(t, t) = \mathbb{E}[X^2(t)] - \mu_x^2(t)$
- It is interpreted as the classic covariance

5.1.5 Auto-Correlation Function

$$\varphi_x(t, s) = \frac{\sigma_x(t, s)}{\sigma_x(t)\sigma_x(s)} = \frac{C_{x,x}(t, s)}{C_{x,x}(t)C_{x,x}(s)} = \frac{\text{autocovariance}}{\text{SD of s times SD of t}}$$

In context of **signal processing** and in **engineering literature** the **autocorrelation function** is denoted as $R_{x,x}(t, s)$ and is defined as:

$$R_{x,x}(t, s) = \mathbb{E}[X(t)X(s)]$$

And it is equivalent to $\sigma_x(t, s)$ only when the mean = 0 and the variance = 1

5.2 Stationary and wide-sense stationary processes

5.2.1 Strongly / Strict-Sense Stationary

A stochastic process is called **Strongly / Strict-Sense Stationary** if:

- All its *statistical proprieties* are **invariant over time**
- for any points t_1, \dots, t_r and any value τ if the two following **joint distributions** are **equivalent**

$$X(t_1), \dots, X(t_r) \equiv X(t_1 + \tau), \dots, X(t_r + \tau)$$

It has the following proprieties:

- $X(t)$ and $X(t + \tau)$ have the *same distribution*, thus same mean, variance, sd
- Since the **joint distribution** of $X(t_1)$ and $X(t_2)$ is invariant respect to its statistical proprieties over time (can be shifted over time with no changes in proprieties) or more shortly **translation invariant** also the **autocovariance** of $X(t)$ must be **translation invariant**
- Same invariance $\forall r \geq 1$

5.2.2 Weakly / Wide-Sense Stationary

A stochastic process $X(t)$ is **Weakly / Wide-Sense Stationary** if the following two conditions holds:

1. The **Mean Function** of $X(t), X(s)$ is *constant*, thus:

$$u(t) \xrightarrow[t \rightarrow \infty]{} \mu$$

2. The **Auto-Covariance Function** of $X(t), C_{xx}(t, s)$ *depends* only on $(s - t) = \tau$, thus:

$$\sigma(t, t + h) \xrightarrow[t \rightarrow \infty]{} \sigma(h) \quad \text{depends only on the distance}$$

5.3 Markov Processes

A stochastic process $X = \{X(t) : t \geq 0\}$ is **Markov** for any $t_1 < t_2 < \dots < t_n < t$ and for any sets (events) A, A_1, \dots, A_n :

$$\mathbb{P}\left[X(t) \in A | X(t_1) \in A_1, \dots, X(t_{n-1}) \in A_{n-1}, X(t_n) \in A_n\right] =$$

$$\mathbb{P}\left[X(t) \in A | X(t_n) \in A_n\right]$$

$$\mathbb{P}\left[\text{future} | \text{past}, \text{present}\right] = \mathbb{P}\left[\text{future} | \text{present}\right]$$

The **future** given the **present** is **independent** from the **past**

Markov Propriety

For a **Markov Process** the *conditional distribution* of $X(t)$ is the **same** under two **different conditions**:

1. Given observations of the process X at several moments in the **past**
2. Given only the **present**, so the latest observation of X

5.3.1 Markov Chain

The **Markov Chain** is a stochastic process with **discrete space** and the **Markov Propriety**.

From now all processes we will study will be **Markov Chain (continuous time)** unless otherwise started.