Probabilistic Methods (PM - 330725)

TOPIC 1: Foundations of Probability

Lecture 2

May, 2025

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Politècnica Superior d'Enginyeria de Manresa

Roadmap

Learning Outcomes

By the end of this topic, you will be able to:

- Grasp foundational concepts like probability spaces, random variables, and distributions.
- Compute and interpret expectation and variance for various probability distributions.
- Use discrete and continuous distributions to model and analyze data.

RECAP

Introduction to Probability and Statistics:

- Probability Spaces, Events, and Axioms
- Introducing Conditional Probability
- Probability calculus
- The multiplication rule
- Independence, also when go wrong!

Bayes' Rule

Key fact: all probabilities are contingent on what we know so far.

When our knowledge changes, our probabilities must change, too.

Bayes' rule tells us how to change them.

- Suppose A is some event we're interested in and B is some new relevant information.
- Bayes' rule tells us how to move from a prior probability, $P(A)$, to a posterior probability that incorporates our knowledge of B.

Bayes' Rule

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

Bayes' rule tells us how to change them.

- $P(A)$ is the prior probability: how probable is , before having seen data B?
- $P(A|B)$ is the posterior probability: how probable is , now that we've seen data B?
- $P(B|A)$ is the likelihood: if were true, how likely is it that we'd see data B?
- $P(B)$ is the marginal probability of B: how likely is it that we'd see data B overall, regardless of whether A is true or not??

Calculating $P(B)$: use the rule of total probability.

Bayes' Rule: a toy example

Imagine a jar with 1024 normal quarters. Into this jar, a friend places a single two-headed quarter (i.e. with heads on both sides). Your friend shakes the jar to mix up the coins. You draw a single coin at random from the jar, and without examining it closely, flip the coin ten times. The coin

comes up heads all ten times.

Are you holding the two-headed quarter, or an ordinary quarter?

- Not just a toy example: in any industry where companies compete strenuously for talent, a lot of time and energy is spent looking for "two-headed quarters"!

Bayes' Rule: a toy example

A real-world version of the two-headed quarter problem:

- Suppose you're in charge of a large trading desk at a major Wall Street bank.
- You have 1025 employees under you, and each one is responsible for managing a portfolio of stocks to make money for your firm and its clients.
- One day, a young trader knocks on your door and confidently asks for a big raise. You ask her to make a case for why she deserves one.

Then, she replies:

> Hey...look at my record: I'm the best trader on your floor. I've been with the company for ten months, and in each of those ten months, my portfolio returns have been in the top half of all the portfolios managed by my peers on the trading floor. If I were just one of those other average Joes, this would be very unlikely. In fact, the probability that an average trader would see above-average results for ten months in a row is only $(1/2)^{10}$, which is less than one chance in a thousand. Since it's unlikely I would be that lucky, I should get a raise.

Bayes' Rule: a toy example

Is the trader lucky, or good? Same math as the big jar of quarters!

- Metaphorically, the trader is claiming to be the two-headed coin (T) in a sea of mediocrity.
- Her data is "D = ten heads in a row": she's performed above average for ten months straight.
- This is admittedly unlikely: $P(D|T) = 1/2^{10} = 1/1024$.
- But excellent performers are probably also rare, so that the prior probability $P(T)$ is pretty small to begin with.

To make an informed decision, you need to know $P(D|T)$: the posterior probability that the trader is an above-average performer, given the data.

Bayes' Rule: a toy example

So let's return to the two-headed quarter example and see how a posterior probability is calculated using Bayes' rule:

$$P(T|D) = \frac{P(T) \cdot P(D|T)}{P(D)}$$

We'll take this equation one piece at a time.

Bayes' Rule: a toy example

$$P(T|D) = \frac{P(T) \cdot P(D|T)}{P(D)}$$

$P(T)$ is the prior probability that you are holding the two-headed quarter.

- There are 1025 quarters in the jar: 1024 ordinary ones, and one two-headed quarter.
- Assuming that your friend mixed the coins in the jar well enough, then you are just as likely to draw one coin as another.
- So $P(T)$ must be 1/1025.

Next, what about $P(T|D)$, the likelihood of flipping ten heads in a row, given that you chose the two-headed quarter?

- Clearly this is 1.
- If the quarter has two heads, there is no possibility of seeing anything else.

Bayes' Rule: a toy example

$$P(T|D) = \frac{P(T) \cdot P(D|T)}{P(D)}$$

Finally, what about , the marginal probability of flipping ten heads in a row? Use the rule of total probability:

$$P(D) = P(T) \cdot P(T|D) + P(notT) \cdot P(D|notT)$$

- $P(T) = 1/1025m$ so $P(notT)$ is $1024/1025$.
- $P(D|T) = 1$.
- $P(D|notT)$ is the probability of a ten-heads "winning streak":

$$P(D|notT) = \binom{1}{2}^{10} = \frac{1}{1024}$$

Bayes' Rule: a toy example

We can now put all these pieces together:

$$P(D|T) = \frac{P(T) \cdot P(D|T)}{P(T) \cdot P(D|T) + P(notT) \cdot P(D|notT)} = \frac{\frac{1}{1025} \cdot 1}{\frac{1}{1025} \cdot 1 + \frac{1024}{1025} \cdot \frac{1}{1024}} = \frac{1}{2}$$

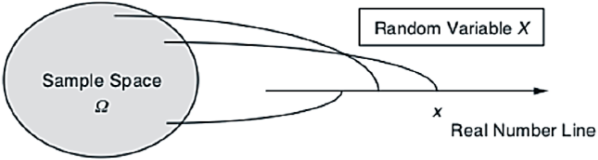
- There is only a 50% chance that you are holding the two-headed coin.
- Yes, flipping ten heads in a row with a normal coin is very unlikely (low likelihood).
- But so is drawing the one two-headed coin from a jar of 1024 normal coins! (Low prior probability.)

Random variables

Defining a Random Variable A random variable X is a measurable function

from the sample space Ω to the real numbers $\Re$. Essentially, it assigns a numerical value to each outcome in Ω. Suppose we have an uncertain

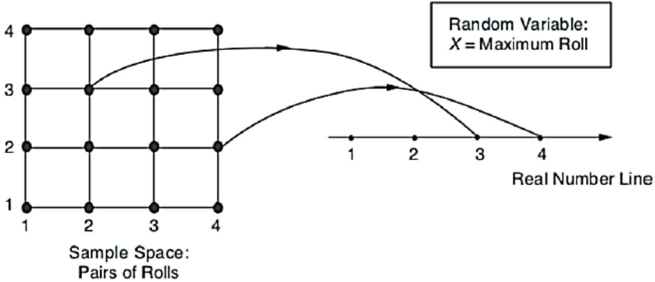
outcome with sample space Ω.
A random variable X is a real-valued function of the uncertain outcome. That is, it maps each element $\omega \in \Omega$ to a real number.



Simply speaking!: a random variable is a numerical summary of some uncertain outcome.

Example 1: rolling two dice

Suppose you roll two dice (an uncertain outcome). An example of a random variable is the maximum of the two rolls:



Random variables: big-picture

Starting with some random outcome with sample space Ω

- A random variable is a real-valued function (i.e. numerical summary) of the outcome ω. Often this outcome fades into the background, but it's always there lurking.
- We describe the behavior of a random variable in terms of its probability distribution : a set of possible outcomes for the random variable, together with their probabilities.
- We can associate with each random variable certain "averages"or "moments"of interest (e.g. mean, variance).
- A function of a random variable defines another random variable.

Discrete random variables

- A random variable is discrete if its range, i.e. the set of values it can take, is a finite or countably infinite set.
- Both our examples (dice, airline no-shows) were discrete random variables: you can count the outcomes on your fingers and toes. Something that can take on a continuous range of values (e.g. temperature, speed) is called a continuous random variable.
- For now, we will focus exclusively on discrete random variables. (The math and notation for continuous random variables is more fiddly, but the concepts are the same.)

Resuming

- **Discrete:** Takes values in a countable set, e.g. $0, 1, 2, \ldots$
- **Continuous:** Takes values in an interval of $\Re$ (uncountably many values).

Probability mass function

Suppose that is a discrete random variable whose possible outcomes are some set $\mathcal{X}$. Then the probability mass function of is

$$p_X(x) = P(\{X \in S\})$$

We always use for X the random variable, and for some possible outcome.

Facts about PMFs:

- $0 \leq p_X(x) \leq 1$
- $\sum_{x \in \mathcal{X}} p_X(x) = 1$
- For any set of numbers S, $P(\{X \in S\}) = \sum_{x \in \mathcal{X}} p_X(x)$

Probability mass functions: an example

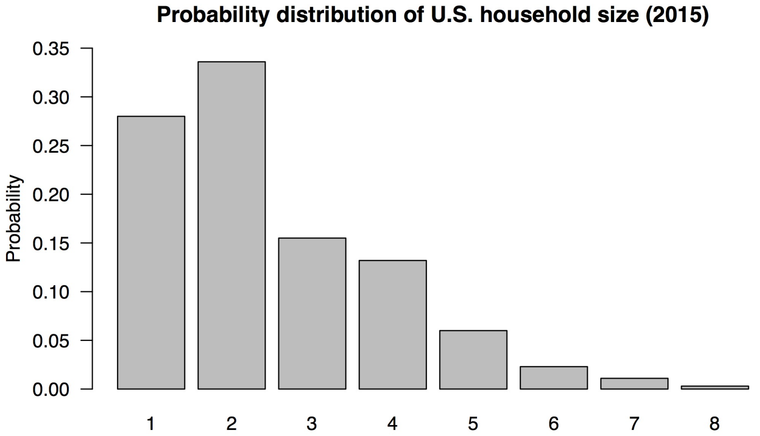
This probability mass function provides a complete representation of your uncertainty in this situation. It has all the key features of any probability distribution:

- An uncertain outcome: the composition of the family next door.
- A numerical function of that uncertain outcome: how large the family is.
- Finally, there are probabilities for each possible value of the random variable (here provided in a look-up table or bar graph).

Most probability distributions won't be this simple, but they all have these same three features.

Expected value

When you knock on the door, how many people do you "expect"to be living next door?



Expected value

The expected value or expectation of a random variable is the weighted average of the possible outcomes, where the weight on each outcome is its probability.

Size x	1	2	3	4	5	6	7	8
Probability, P(X = x)	0.280	0.336	0.155	0.132	0.060	0.023	0.011	0.003

$$E(X) = (0.280) \cdot 1 + (0.336) \cdot 2 + \ldots + (0.011) \cdot 7 + (0.003) \cdot 8 \approx 2.5$$

- The more likely numbers (e.g. 1 and 2) get higher weights than,
- while the unlikely numbers (e.g. 7 and 8) get lower weights.

Expected value

This differs from the ordinary average! For our family-size example this would be:

$$E(X) = \frac{1}{8} \cdot 1 + \frac{1}{8} \cdot 2 + \ldots + \frac{1}{8} \cdot 7 + \frac{1}{8} \cdot 8 \approx 4.5$$

- Here, the weight on each number in the sample space is $1/8 = 0.125$, since there are 8 numbers
- This is not the expected value; it give each number in the sample space an equal weight, ignoring the fact that these numbers have different probabilities.

Expected value

This example conveys something important about expected values. Even if the world is black and white, an expected value is often grey:

- the "expected"American household size is 2.5 people
- a baseball player "expects"to get 0.25 hits per at bat.
- you "expect"to get 0.5 heads per coin flip.
- you "expect"that your newborn child will have one 0.97 ovaries (since 100/206 newborns are female).

Even if the underlying outcomes are all whole numbers, the expected value doesn't have to be.

Expected value: definition

Suppose that the possible outcomes for a random variable X are the numbers $x_1, x_2, \ldots, x_N$ with probabilities $P(X = x_i)$.

The formal definition for the expected value of X is:

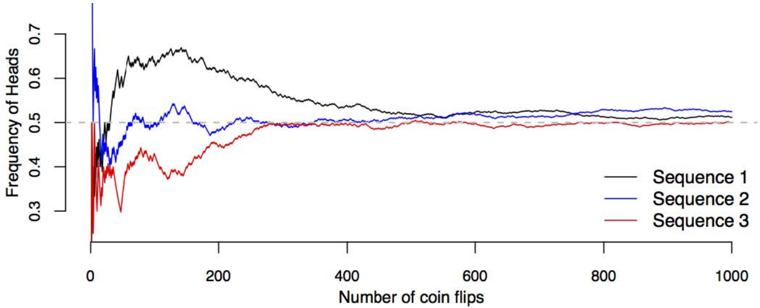
$$E(X) = \sum_{i=1}^{N} P(X = x_i) \cdot x_i = \sum_{x \in \mathcal{X}} p_X(x) \cdot x$$

Reminder!

This measures the çenter"or mean of the probability distribution.

An aside: long-run averages

You may have heard the expected value explained intuitively in terms of a limiting long-run average:



Three simulated long-run sequences of coin flips.

An aside: long-run averages

You may have heard the expected value explained intuitively in terms of a limiting long-run average:

Suppose we were to observe a large number of random variables with the same probability distribution P. The expected value of P is the limiting long-run average of the observed outcomes.

Spoiler!:

This is 100% true! But it is not the definition of the expected value.

- Rather, it follows from a really cool theorem called the Law of Large Numbers.

Variance: definition

A related concept is the variance, which measures the dispersion or spread of a probability distribution.

$$var(X) = E(\{X - E(X)\}^2)$$

- The standard deviation of a probability distribution is

$$\sigma = sd(X) = \sqrt{var(X)}$$

- The standard deviation is more interpretable than the variance, because it has the same units (dollars, miles, etc.) as the random variable itself.

Again suppose that there are possible outcome $x_1, x_2, \ldots, x_N$.

Then by just applying the definition of expected value, we see that:

$$var(X) = E(\{X - E(X)\}^2) = \sum_{i=1}^{N} P(X = x_i) \cdot (x_i - \mu)^2$$

where $\mu = E(X)$ is the expected value of the random variable.

A useful identity for the variance of a random variable is

$$var(X) = E(\{X - E(X)\}^2) = E(X^2) - E(X)^2$$

Maybe not super intuitive, but it is sometimes useful for calculating stuff.

Building probability models

A probability model is a stylized description of a real-world system in terms of random variables. To build one:

1. Identify the uncertain outcome of interest (e.g. a soccer game between Arsenal and Manchester United) and corresponding sample space Ω (e.g. all possible soccer scores).
2. Identify the random variables associated with that uncertain outcome, e.g. X_A = the number of goals scored by Arsenal, and X_M = the number of goals scored by Man U. (Remember: a random variable is just a numerical summary of an uncertain outcome.)
3. Finally, specify a probability distribution that lets us calculate probabilities associated with each random variable.

Thus instead of building probability distributions from scratch, we will rely on a simplification called a parametric probability model.

Parametric models

A parametric probability distribution, or parametric model, is a probability distribution that can be completely described using a relatively small set of numbers.

These numbers are called the parameters of the distribution. Lots of parametric models have been invented for specific purposes:

- normal, binomial, Poisson, t-student, chi-squared, Weibull, etc.
- A large part of getting better at probability modeling is to learn about these existing parametric models: what they are, and when they're appropriate.

Parametric models

Suppose X is a discrete random variable with possible outcomes $x \in \mathcal{X}$.
In a parametric model, the PMF of takes the form

$$P(X = x) = f(x; \theta)$$

- θ where is the parameter (or parameters) of the model.
- The parameter θ completely specifies the PMF
- Contrast this with the family-size example, where we had to write out the PMF explicitly for each possible outcome.

Continuous Distribution Function

Up to now we've only been dealing with discrete random variables that are characterized by a PMFs.

But what about a random variable like:

- X = Apple's stock price tomorrow?
- X = speed of the next pitch thrown by Justin Verlander?
- X = blood pressure of a randomly sampled participant in a clinical trial of a new drug?

These outcomes cannot naturally be restricted to a finite or countable set, and they don't have PMFs. To describe these random variables, we need some more general concepts.

Continuous Distribution Function

The cumulative distribution function, or CDF, is defined as:

$$F_X(x) = P(X \leq x)$$

Some Facts:

- All random variables have a CDF.
- The CDF completely characterizes the random variable: if X and Y have the same CDF, then for all sets S, $P(X \in S) = P(Y \in S)$.
- If this holds, we say that X and Y are equal in distribution. This doesn't mean they're identical! It just means that all probability statements about and will be identical.

Properties of CDFs

All CDFs satisfy the following properties:

1. F is non-decreasing: if $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$

2. F is bottoms out at 0 and tops out at 1:
   - $lim_{x \to -\infty} F(x) = 0$
   - $lim_{x \to \infty} F(x) = 1$

3. F is right-continuous, i.e.

$$F(x) = lim_{y \downarrow x} F(y)$$

Note: $lim_{y \downarrow x}$ means l̇limit as approaches from above."

Intuitively, a continuous random variable is one that has no "jumpsïn its CDF. More formally, we say that is a continuous random variable if there exists a function such that:

- $f(x) \geq 0$
- $\int_{\mathcal{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$
- For every interval $S = (a, b)$,

$$P(X \in S) = P(a \leq X \leq b) = \int_a^b f(x)dx$$

The PDF/CDF relationship

Note that, by the definition of the CDF and PDF, we have the following relationship for a continuous random variable:

$$F_X(x) = P(X \in (-\infty, x)) = \int_{-\infty}^{x} f_X(x)dx$$

Remember the Fundamental Theorem of Calculus! This relationship says that the PDF is the derivative of the CDF:

$$f(x) = F'(x)$$

at all points where is differentiable!

Expected value and Variance

Remember expected value and variance for discrete random variables. Suppose that takes the values $x_1, x_2, \ldots, x_N$. Then

$$\mu = E(X) = \sum_{i=1}^{N} x_i \cdot P(X = x_i)$$

and,

$$\sigma^2 = var(X) = \sum_{i=1}^{N} (x_i - \mu)^2 \cdot P(X = x_i)$$

In the continuous case, the sum becomes an integral:

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

and,

$$\sigma^2 = var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Warning!

Continuous random variables and PDFs can be confusing.

- If is continuous, then $P(X = x) = 0$ for every point. Only sets with nonzero length have positive probability. If this seems weird, blame it on the real number system.
- The PDF doesn't give you probabilities directly. That is, $f_X(x) \neq P(X = x)$. This only holds for the PMF of a discrete random variable.
- Unlike a PMF, a PDF can be larger than 1. For example, say $f(x) = 3$ for $0 \leq x \leq 1/3$, and $f(x) = 0$ otherwise. This is a well-defined PDF since $\int_{\mathcal{R}} f(x)dx = 1$.

Here a some useful facts about CDFs

- $P(a \leq X \leq b) = F_X(b) - F_X(a)$
- $P(X \geq x) = 1 - F_X(x)$.
- If X is continuous, then $P(a \leq X \leq b) = P(a \leq X \leq b) = P(a \leq X \leq b) = P(a \leq X \leq b)$

(Including/excluding endpoints makes no difference.)

Probability distributions: continuous, discrete

- A probability distribution P determines the probability of different outcomes of a random variable
- Probability distributions for:
  - discrete random variables which have finite or countable sets of possible outcome values (e.g., dice, coins, birthdays);
  - continuous random variables which have infinite sets of possible outcome values (e.g., temperature, length).
- The corresponding probability distributions: continuous, discrete.

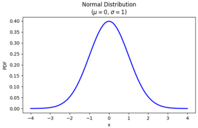
(Remark. Actually, there are distributions which are neither continuous nor discrete.)

Probability density functions

Examples of the probability density p of some continuous distributions (realized also in R with some default parameter values):

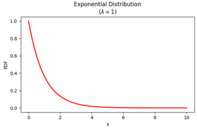
Normal distribution `norm` with parameters μ (mean = 0) and σ (sd = 1)

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right), \quad x \in \mathbb{R}.$$



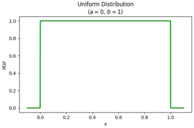
Exponential distribution `exp` with parameter λ ($\lambda = 1$)

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$



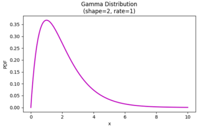
Uniform distribution `unif` with parameters min $= a$ and max $= b$

$$p(x) = \frac{1}{b-a}, \quad a \le x \le b.$$



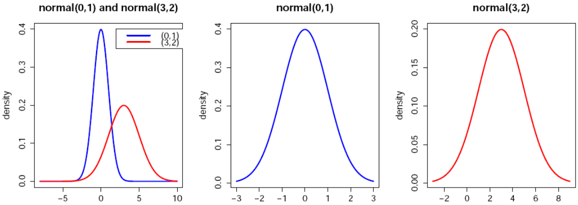
Gamma distribution `gamma` with parameters *shape* and *rate* = 1

$$p(x) = \frac{\text{rate}^{\text{shape}}}{\Gamma(\text{shape})}\, x^{\text{shape}-1} \exp(-\text{rate}\, x).$$



All normal distributions are similar

Two normal distributions with different parameters μ and σ are still similar.



All normal distributions have the same bell shape. Recall its density function:

$$p_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}.$$

Probability mass functions for some discrete distributions

If a random variable X has a distribution with the density $p(x)$, then

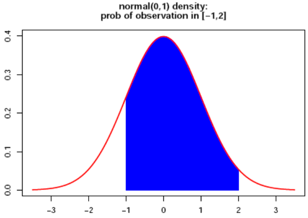
$$P\big(X \in I\big) = \int_I p(x)dx \qquad \text{for (almost) any interval} \quad I \subseteq \mathbb{R}.$$

In other words, the probability to have an outcome in some interval I is the area under the density function $p(x)$ over that interval.

Example. For $X \sim N(0,1)$,

$$P(-1 \leq X \leq 2) = P\big(X \in [-1,2]\big)$$
$$= \int_{-1}^{2} p(x)dx = \int_{-1}^{2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.82.$$

In events for continuous distributions:
$<$ or $\leq$ ($>$ or $\geq$) does not matter.



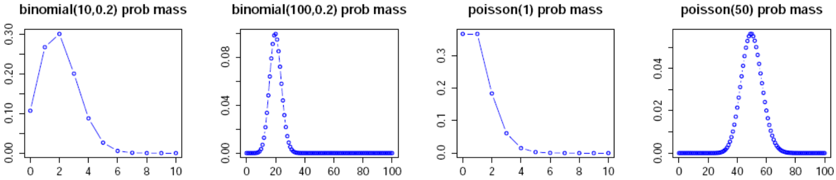
For discrete distributions we have a probability mass function p

$$p(x) = P(X = x).$$

The probability to have an outcome in some set A is the sum

$$P\big(X \in A\big) = \sum_{x \in A} p(x).$$

Examples of discrete distributions are binomial and Poisson.



Probability mass functions for some discrete distributions

- **Binomial distribution** `binom` with parameters n (`size`) and p (`prob`)

$$p(x) \;=\; \frac{n!}{x!\,(n-x)!}\, p^x\,(1-p)^{\,n-x}.$$

- **Poisson distribution** `pois` with parameter λ (`lambda`)

$$p(x) \;=\; \frac{\lambda^x}{x!}\, e^{-\lambda}.$$

Cumulative distribution/probability function

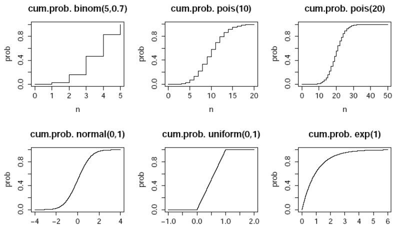
The cumulative distribution function (CDF) (sometimes also called cumulative probability function) of a random variable X is

$$F(u) = P(X \leq u) = \texttt{pdist(u,par)} \qquad \text{(continuous and discrete).}$$

Continuous distributions: $F(u) = \int_{-\infty}^{u} p(x)\, dx$

Discrete distributions: $F(u) = \sum_{x \leq u} p(x)$. Any other probability can be computed via $F(u)$. For any $a \leq b$,

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

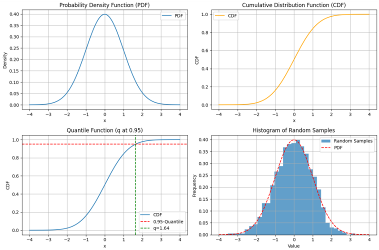


Python and R commands for distributions

- In R or Python (SciPy), there are a number of continuous and discrete distributions `dist` with parameters `par`.
- Let $p(x)$ denote the density for continuous distributions and the probability mass function for discrete distributions.
- `ddist(x, par)` computes $p(x)$ (i.e., either density or mass function), `dist.pdf(x, *par)` (continuous) & `dist.pmf(x, *par)` (discrete)
- `pdist(u, par)`, `dist.cdf(u, *par)` computes the CDF $F(u) = P(X \leq u)$,
- `qdist(a, par)`, `dist.ppf(a, *par)` ($a \in [0, 1]$) computes the value q such that
`pdist(q, par) = a`, this is the a-quantile. The α-quantile q_α is such number that $P(X \leq q_\alpha) = \alpha$.
- `rdist(size, par)`, `dist.rvs(*par, size=size)` yields a random sample from `dist` with parameter `par` of size `size`.

Python and R commands for distributions

- **Probability Density Function (PDF)**: It shows the likelihood of different outcomes for a continuous random variable (in this example, a Normal distribution).
- **Cumulative Distribution Function (CDF)**: It shows the probability that the variable takes a value less than or equal to x.
- **Quantile Function**: Illustrates the 0.95 quantile, highlighting the point below which 95% of the distribution's values fall.
- **Histogram of Random Samples**: Depicts a histogram from random samples generated from the distribution, overlaid by the PDF to show the sample distribution aligns with the theoretical distribution.



Example_Lecture2.ipynb

```python
from scipy.stats import norm, binom, poisson

# Normal distribution probabilities: P(-1 < X <= 2)
prob_normal = norm.cdf(2, loc=0, scale=1) - norm.cdf(-1, loc=0, scale=1)
print(f"P(-1<X<=2) = {prob_normal:.6f}")
```

```
P(-1<X<=2) = 0.818595
```

```python
# Generating 4 standard normal random samples
samples_normal = norm.rvs(size=4, loc=0, scale=1)
print("4 standard normal samples:", samples_normal)
```

```
4 standard normal samples: [ 0.37288429  0.2079086  -0.98259379 -0.36435254]
```

```python
# Binomial distribution probability: P(X = 1) with size=5 and prob=0.2
prob_binom_1 = binom.pmf(1, n=5, p=0.2)
print(f"P(X=1) Binomial(n=5,p=0.2): {prob_binom_1:.4f}")
```

```
P(X=1) Binomial(n=5,p=0.2): 0.4096
```

```python
# Binomial distribution cumulative probability: P(X <= 1)
prob_binom_cdf_1 = binom.cdf(1, n=5, p=0.2)
print(f"P(X<=1) Binomial(n=5,p=0.2): {prob_binom_cdf_1:.5f}")
```

```
P(X<=1) Binomial(n=5,p=0.2): 0.73728
```

```python
# Checking the cumulative binomial probability explicitly (P(X=0) + P(X=1))
prob_binom_explicit = binom.pmf(0, 5, 0.2) + binom.pmf(1, 5, 0.2)
print(f"P(X=0)+P(X=1) explicitly: {prob_binom_explicit:.5f}")
```

```
P(X=0)+P(X=1) explicitly: 0.73728
```

```python
# Generate 3 Poisson random samples with lambda=5
samples_poisson = poisson.rvs(mu=5, size=3)
print("3 Poisson samples (lambda=5):", samples_poisson)
```

```
3 Poisson samples (lambda=5): [5 3 3]
```

Recap: Expectation

- The expectation or mean $E(X)$ of a random variable X with probability distribution P is a location parameter of distribution P.
- For discrete random variable: $E(X) = \sum_x xp(x)$.
- For continuous random variable: $E(X) = \int xp(x)dx$.

Examples

Throwing a dice: $E(X) = \sum_x xp(x) = 1 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3\frac{1}{2}$.

Normal distribution:

$$E(X) = \int xp(x)dx = \int x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \cdots = \mu.$$

Recap: Variance and standard deviation

- The variance of a probability distribution is a scale (or spread) parameter.
- For discrete random variable: $Var(X) = \sum_x (x - E(X))^2 p(x)$.
- For continuous random variable:
  $Var(X) = \int (x - E(X))^2 p(x) dx$.
- Definition: the standard deviation σ is the square root of the variance $\sigma = \sqrt{Var(X)}$.

Examples

Throwing a dice:

$$Var(X) = \sum_x (x - 3\tfrac{1}{2})^2 p(x) = (1 - 3\tfrac{1}{2})^2 \times \tfrac{1}{6} + \cdots + (6 - 3\tfrac{1}{2})^2 \times \tfrac{1}{6} = 2.92.$$

Normal distribution:

$$Var(X) = \int (x - \mu)^2 p(x) dx = \int (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \cdots = \sigma^2.$$

Expectation and variance for some distributions

	expectation	**variance**
uniform$(a,\ b)$	$(a+b)/2$	$(b-a)^2/12$
normal$(\mu,\ \sigma^2)$	μ	σ^2
exponential(λ)	$1/\lambda$	$1/\lambda^2$
binomial$(n,\ p)$	np	$np(1-p)$
poisson(λ)	λ	λ

Wrapping up

Today we discussed:

- Probability and statistics provide the mathematical language and tools for dealing with uncertainty in diverse fieldsâĂŤfrom physics and engineering to finance, data science, biology, and social sciences.
- It involves:
  1. from Conditional Probability to Bayes's Rule.
  2. Random Variables in both discrete and continuous it's important to distinguish
  3. Probability mass function&Continuous Distribution Function, and both Expected value and Variance

References

Online Tutorials, Courses, and other books

- Chapte 2 - Ross, S. Introduction to probability models. 13th edition. Amsterdam: Academic Press, 2023. ISBN 9780443187612.
- MIT OpenCourseWare:
  Probabilistic Systems Analysis and Applied Probability
- Other books:
  All of Statistics: A Concise Course in Statistical Inference

Thank you very much!

ANY QUESTIONS OR COMMENTS?