Week 5: Continuous random variables

Armenak Petrosyan

Last time

mgf of Poisson distribution

- From the definition of mgf:

$$M(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t-1)}.$$

- Therefore

$$M'(t) = \lambda e^t e^{\lambda(e^t-1)}$$
$$M''(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)}$$

- Finally

$$E[X] = M'(0) = \lambda \quad \text{(as expected)}$$
$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = M''(0) - \lambda^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

Exercise 1

Problem (2.6-5 from the textbook)

Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume a Poisson distribution, find the probability of at most one flaw appearing in 225 square feet.

Solution

- *If the average on 150 sq ft is 1 then the average on 225 sq ft will be*

$$\lambda = \frac{225}{150} = \frac{3}{2}.$$

- *We want to find*

$$f(0) + f(1) = \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} = \frac{5}{2}e^{-\frac{3}{2}} \approx 0.5578.$$

Tuesday class

Continuous random variables

So far X was a discrete random variable:

- $\text{Range}(X)$ was a discrete set.
- The pmf and cdf had all the information we needed about X.

$$\boxed{F(x) = \sum_{y \in \text{Range}(X) \text{ and } y \leq x} f(y)}$$

$$\boxed{f(x) = F(x+1) - F(x).}$$

Definition (Cumulative distribution function)

Let X be any random variable (does not have to be discrete). Define the cdf of X as before

$$F(x) = P(X \leq x).$$

Turns out, for some non-discrete random variables X,

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}.$$

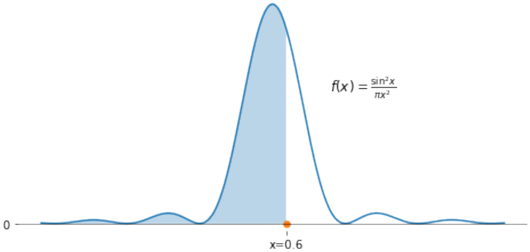
For non-discrete random variables, the pmf does not tell us much about the random variable.

Definition (Continuous random variable, pdf)

X is called a **continuous random variable** if there exists a function $f(x)$ such that

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y.$$

$f(x)$ is called the **probability density function** or **pdf** of continuous random variable X.



For continuous r.v., to draw cdf, we normally draw the graph of pdf and shade the area left of x.

- It is called continuous, because the cdf $F(x)$ is a continuous function. In fact, it has a derivative everywhere, except “a few” points .
- From the fundamental theorem of calculus,

$$\boxed{f(x) = F'(x) \text{ for all } x \in \mathbb{R}}$$

- For any $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

pmf vs pdf

- pmf is a probability and is always ≤ 1.
- pdf is the concentration of probability around x and can be any non-negative number

$$f(x) = \lim_{\delta \to 0} \frac{F(x+\delta) - F(x)}{\delta} = \lim_{\delta \to 0} \frac{P(x < X \leq x + \delta)}{\delta}.$$

Example: uniform distribution

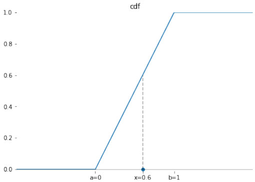
- Consider the experiment of aimlessly throwing a point on the interval $S = [a, b]$.
- Let X be the value of where the point lands.
- $\text{Range}(X) = [a, b]$.
- For any $[c, d] \subset [a, b]$, $P(X \in [c, d]) = \frac{d-c}{b-a}$.
- Therefor, for $x \in [a, b]$,

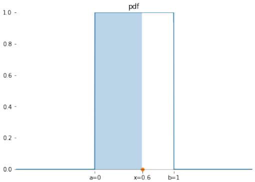
$$P(X \leq x) = P(X \in [a, x]) = \frac{x-a}{b-a}.$$

Example: uniform distribution

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$

$$f(x) = F'(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$





Definition

A continuous random variable is called uniform if $f(x)$ is constant on $\mathrm{Range}(X)$.

Theorem

If $f(x)$ is a pdf of a continuous random variable then

1. $f(x) \geq 0$ *for every* $x \in \mathbb{R}$.
2. $\int\limits_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$.

Proof.

1. $F(x)$ is non-decreasing therefore $F'(x) \geq 0$.
2. Notice that
$$\int\limits_{\mathbb{R}} f(x)\,\mathrm{d}x = P(S) = 1.$$

$\square$

Expected value of a continuous r.v.

Definition (Expected value)

Let X be a continuous random variable. The **expected value** or the **mean** of X is the number

$$E[X] = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x,$$

assuming

$$\int_{\mathbb{R}} |x| f(x)\,\mathrm{d}x < \infty.$$

If the later condition holds, we say the the **expected value of X exists**.

Example

Let X have uniform distribution on $[a, b]$. Then

$$E[X] = \int_a^b x \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b = \frac{a+b}{2}.$$

Example

$$f(x) = \begin{cases} \frac{1}{3} & 0 \le x \le 1 \\ \frac{2}{3} & 1 \le x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

- This is a pdf:

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x + \int_1^2 f(x)\,\mathrm{d}x = \frac{1}{3} + \frac{2}{3} = 1.$$

- Expected value:

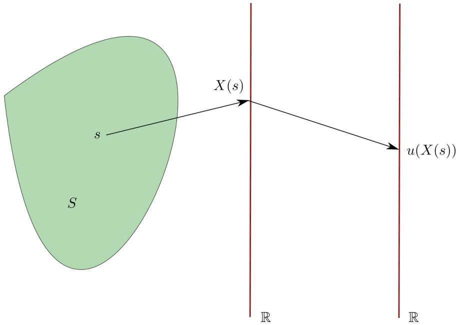
$$\begin{aligned} E[X] = \int_{\mathbb{R}} x f(x)\,\mathrm{d}x &= \int_0^1 x\frac{1}{3}\,\mathrm{d}x + \int_1^2 x\frac{2}{3}\,\mathrm{d}x \\ &= [1^2 - 0^2]\frac{1}{3} + [2^2 - 1^2]\frac{2}{3} \\ &= \frac{1}{3} + 2 = \frac{5}{3}. \end{aligned}$$

Change of a random variable

- Let $X : S \to \mathbb{R}$ be a continuous random variable on the set of outcomes S.
- Let $u : \mathbb{R} \to \mathbb{R}$ be any function (e.g. $u(x) = x^2$).
- Then the composition function

$$Y = u(X), \quad u(X) : S \to \mathbb{R}$$

will be a new random variable.



Theorem

Let X be a continuous random variable with pdf $f_X(x)$ and $u : \mathbb{R} \to \mathbb{R}$ be a function. If

$$\int_{\mathbb{R}} |u(x)| f_X(x)\, \mathrm{d}x < \infty$$

then the expected value of $Y = u(X)$ exists and

$$E[Y] = \int_{\mathbb{R}} u(x) f(x)\, \mathrm{d}x.$$

Proof (sketch)

- Let us do in the simple case when u is differentiatable and strictly increasing.
- Notice that

$$F_Y(y) = P[Y \leq y] = P[u(X) \leq y] = P[X \leq u^{-1}(y)] = F_X(u^{-1}(y)).$$

- Using the chain rule

$$f_Y(y) = F_Y'(y) = f_X(u^{-1}(y)) \left(u^{-1}(y)\right)'.$$

Proof (cont.)

- Then

$$E[Y] = \int_{\mathbb{R}} y f_Y(y)\,\mathrm{d}y$$

(after change of variable $y = u(x)$)

$$= \int_{\mathbb{R}} u(x) f_Y(u(x)) u'(x)\,\mathrm{d}x$$

(after insterting the value for f_Y)

$$= \int_{\mathbb{R}} u(x) f_X(x)\,\mathrm{d}x.$$

$\square$

Variance, moments and mgf of continuous r.v.

Let $\mu = E[X]$. We define

- **Variance**

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \int_{\mathbb{R}} (x-\mu)^2 f_X(x)\,\mathrm{d}x.$$

- **Standard deviation**

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

- **r-th moment**

$$E[X^r] = \int_{\mathbb{R}} x^r f_X(x)\,\mathrm{d}x.$$

- **Moment generating function (mgf)**

$$M(t) = \int_{\mathbb{R}} e^{xt} f_X(x)\,\mathrm{d}x, \quad -h < t < h.$$

Again, we have

$$\mathrm{Var}(X) = E[X^2] - E[X]^2, \quad E[X] = M'(0), \quad E[X^2] = M''(0).$$

Exercise 2

Problem (3.1-11 in the textbook)

The pdf of Y is $g(y) = c/y^3$, $1 < y < \infty$

(a) *Calculate the value of d so that g(y) is a pdf.*

(b) *Find $E(Y)$.*

(c) *Show that $\mathrm{Var}(Y)$ is not finite.*

Solution

(a) $1 = \int\limits_{\mathbb{R}} g(y)\,\mathrm{d}y = \int\limits_{1}^{\infty} \frac{c\,\mathrm{d}y}{y^3} = -\frac{c}{2y^2}\Big|_1^{\infty} = \frac{c}{2} \Rightarrow c = 2.$

(b) $E[Y] = \int\limits_{\mathbb{R}} y g(y)\,\mathrm{d}y = \int\limits_{1}^{\infty} y\frac{2}{y^3} = -\frac{2}{y}\Big|_1^{\infty} = 2.$

(c)
- $E[Y^2] = \int\limits_{\mathbb{R}} y^2 g(y)\,\mathrm{d}y = \int\limits_{1}^{\infty} y^2\frac{2}{y^3} = \log(y)|_1^{\infty} = \infty.$
- $\mathrm{Var}(Y) = E[Y^2] - E[Y]^2 = \infty.$