5: Continuous Random Variables

PSTAT 120A: Summer 2022

Ethan P. Marzban
June 30, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete distributions

Continuous Random Variables

Recap of RV's

- Let's recap what we know about random variables.
- They map from Ω to $\mathbb{R}$
- The state space is defined to be the image of Ω, and we classify random variables based on the cardinality of their state space
- One fact that I didn't explicitly mention is that

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$$

Continuous Random Variable

- The construction of continuous random variables is a bit funky.
- We actually *start* with the fourth point above: namely, if $F_X(x) := \mathbb{P}(X \leq x)$ then $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$.
- Remember that

$$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$$

- So, if we have a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ then we can construct $F_X(x)$ very easily.
- It turns out that, under certain conditions, we have the existence of a function $f_X(x)$ that obeys the following key property:

$$\int_a^b f_X(x) \, \mathrm{d}x = F_X(b) - F_X(a)$$

- It further turns out that this function $f_X(x)$ must obey two properties:
  1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
  2. $\int_{\mathbb{R}} f_X(x) \, \mathrm{d}x = 1$
- Such a function $f_X(x)$ is called a **probability density function** (p.d.f.).
  - If it helps, you can think of a p.d.f. as a continuous analog of the p.m.f., but **be careful**; $f_X(x)$ does *NOT* represent a probability, whereas $p_X(k)$ does. The p.d.f. is a purely mathematical construction.

Properties

- $\mathbb{E}[X] := \int_{\mathbb{R}} x f_X(x)\ \mathrm{d}x$
- $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f_X(x)\ \mathrm{d}x$
- $F_X(x) = \int_{-\infty}^{x} f_X(t)\ \mathrm{d}t \implies f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x)$
- $\mathrm{Var}(X) := \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\} = \mathbb{E}[X^2] - [\mathbb{E}(X)]^2$

Comparison of Discrete and Continuous Random Variables

Discrete

probability **mass** function (p.m.f.)

$$p_X(x) := \mathbb{P}(X = x)$$

$$\forall x \; 0 \leq p_X(x) \leq 1$$

$$\sum_{\text{all } x} p_X(x) = 1$$

Continuous

probability **density** function (p.d.f.)

$$f_X(x)$$

$$\forall x \; f(x) \geq 0$$

$$\int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = 1$$

Cumulative Distribution Function (C.D.F.)

$$F_X(x) := \mathbb{P}(X \leq x)$$

Discrete

$$F_X(x) = \sum_{y \leq x} p_X(y)$$

Continuous

$$F_X(x) = \int_{-\infty}^{x} f_X(y) \, \mathrm{d}y$$

Comparison of Discrete and Continuous Random Variables

Expected Value

$$\mathbb{E}(X) = \mu_X$$

Discrete

$$\mathbb{E}(X) = \sum_{\text{all } x} x p_X(x)$$

$$\mathbb{E}[g(X)] = \sum_{\text{all } x} g(x) p_X(x)$$

Continuous

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\ \mathrm{d}x$$

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\ \mathrm{d}x$$

Variance

$$\mathrm{Var}(X) = \sigma_X^2 = \mathbb{E}\left[(X - \mu_X)^2\right] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Discrete

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{\text{all } x} (x - \mu_X)^2 p_X(x) \\ &= \sum_{\text{all } x} x^2 p(x) - [\mathbb{E}(X)]^2\end{aligned}$$

Continuous

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\ \mathrm{d}x \\ &= \left[\int_{-\infty} x^2 f_X(x)\ \mathrm{d}x\right] - [\mathbb{E}(X)]^2\end{aligned}$$

Example (Chalkboard)

Suppose X is a random variable that has p.d.f. given by

$$f_X(x) = \begin{cases} cx & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

where $c > 0$ is an as-of-yet undetermined constant.

(a) What is the value of c?

(b) Compiute $\mathbb{P}(X = 0.5)$.

(c) Compute $\mathbb{P}(X \in [0.25, 0.75])$

(d) Compute $\mathbb{E}[X]$

(e) Compute $\mathbb{E}\left[\frac{1}{X}\right]$

Constructing a P.D.F.

- Let's go through an example of how to construct a p.d.f.
- First, let's start off with a new probability measure: if $\Omega = [a, b]$ it turns out that

$$\mathbb{P}(A) = \frac{\text{length}(A)}{b - a}$$

  is in fact a valid probability measure.

- Then, if we take the probability space $([a, b], \mathcal{F}, \mathbb{P})$ where $\mathbb{P}$ is defined as above, and if we have a random variable $X : [a, b] \to \mathbb{R}$ then

$$F_X(x) = \mathbb{P}([a, x]) = \frac{\text{length}([a, x])}{b - a} = \frac{x - a}{b - a}$$

  provided, of course, that $x \in [a, b]$. Therefore, if $x \in [a, b]$ we have

$$F_X(x) = \frac{x - a}{b - a} \implies f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{x - a}{b - a}\right) = \frac{1}{b - a}$$

- If $x \notin [a, b]$, we can see that $f_X(x) = 0$ and so

$$\boxed{f_X(x) = \begin{cases} \dfrac{1}{b - a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}}$$

- This is in fact the p.d.f. of the so-called **Uniform** distribution: $X \sim \text{Unif}[a, b]$

Continuous Distributions

Continuous Distributions

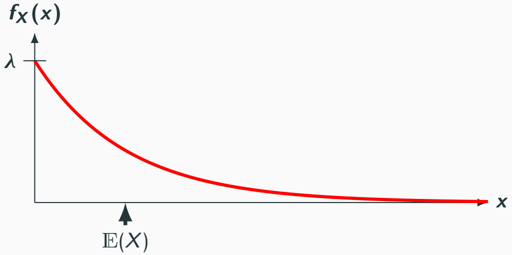
- Just like we had discrete distributions, we also have continuous distributions as well.
- Good news: I won't expect you to derive p.d.f.'s from probability measures like we did on the previous slide. From here on out I'll just give the p.d.f. (or c.d.f.).

Exponential Distribution

- If $X \sim \text{Exp}(\lambda)$, then

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- $\mathbb{E}[X] = \dfrac{1}{\lambda}$
- $\text{Var}(X) = \dfrac{1}{\lambda^2}$
- $F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$



Gamma Distribution

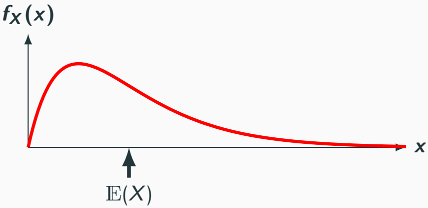
- If $X \sim \text{Gamma}(r, \lambda)$, then

$$f_X(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(r)} \cdot x^{r-1} e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  where $\Gamma(r)$ denotes the **Gamma Function:**

$$\Gamma(r) := \int_0^{\infty} x^{r-1} e^{-x} \, \mathrm{d}x$$

- $\mathbb{E}[X] = \dfrac{r}{\lambda}$
- $\text{Var}(X) = \dfrac{r}{\lambda^2}$
- Note that the Gamma$(1, \lambda)$ distribution is equivalent to the Exp(λ) distribution



Memorylessness

- It turns out that the Exponential Distribution actually possesses a very special property:

Definition: Memorylessness

A distribution is said to possess the **memorylessness property** (or, equivalently, that the distribution is **memoryless**) if for $s, t > 0$

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s)$$

where X is a random variable that follows the distribution in question.

- Here's one way to interpret memorylessness: say X models the lifetime of an electrical component. The memorylessness property says: given that the component has functioned for t units of time, the conditional probability that it works for an additional s units of time is is the same as the unconditional probability that the original component functions for s units of time.
  - That is, regardless of how long the component has functioned the distribution of the remaining lifetime is the same as the distribution of the original (unconditional) lifetime; the lifetime continually renews itself.

Memorylessness

Theorem

The Exponential Distribution is memoryless.

Proof.

- Let $X \sim \text{Exp}(\lambda)$.
- Then

$$\mathbb{P}(X \geq s) = \int_s^\infty \lambda e^{-\lambda x} \, \mathrm{d}x = e^{-\lambda s}$$

- Additionally,

$$\mathbb{P}(X \geq t + s \mid X \geq t) = \frac{\mathbb{P}(\{X \geq t + s\} \cap \{X \geq t\})}{\mathbb{P}(X \geq t)}$$

- For the numerator: note that if $X \geq t + s$ then we automatically have $X \geq t$; that is

$$\{X \geq t + s\} \subseteq \{X \geq t\}$$

  and so

$$\mathbb{P}(\{X \geq t + s\} \cap \{X \geq t\}) = \mathbb{P}(X \geq t + s) = \int_{t+s}^\infty \lambda e^{-\lambda x} \, \mathrm{d}x = e^{-\lambda(t+s)}$$

Theorem

The Exponential Distribution is memoryless.

Proof.

- Therefore,

$$\mathbb{P}(X \geq t + s \mid X \geq t) = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = \frac{e^{-\lambda s} \cdot e^{-\lambda t}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(X \geq s)$$

  which completes the proof.

$\square$

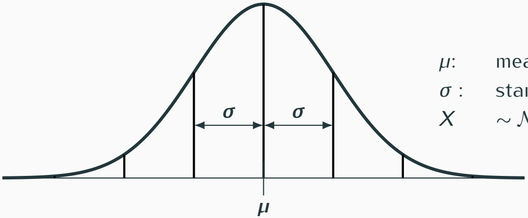
- It can be shown that the Exponential distribution is the *only* memoryless continuous distribution.
- Additionally, it can be shown that the Geometric distribution is the *only* memoryless discrete distribution.

Normal (Gaussian) Distribution

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$f_X(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

- $\mathbb{E}[X] = \mu$
- $\mathrm{Var}(X) = \sigma^2$



μ: mean
σ : standard deviation
X $\sim \mathcal{N}(\mu, \sigma^2)$

Standard Normal Distribution

- If $Z \sim \mathcal{N}(0,1)$, we say X follows ths **standard normal distribution**. Note that:
  - $f_Z(z) =: \phi(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$
  - $\mathbb{E}[X] = 0$, $\mathrm{Var}(X) = 1$
- The c.d.f. of the standard normal distribution arises so frequently, we give it a name: $\Phi(\cdot)$. In other words,

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, \mathrm{d}z$$

- There exist lookup tables for $\Phi(z)$; see the next slide.

Fact: Standardization

- If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $Z := \left(\frac{X-\mu}{\sigma}\right) \sim \mathcal{N}(0,1)$. The act of subtracting the mean and dividing by the standard deviation is called **standardization.**
- If $Z \sim \mathcal{N}(0,1)$ and if $\sigma > 0$, then $X := (\sigma Z + \mu) \sim \mathcal{N}(\mu, \sigma^2)$

Example

If $X \sim \mathcal{N}(1,4)$, compute $\mathbb{P}(X \geq 2)$. **Leave your answer in terms of Φ, the standard normal c.d.f.**

$$\mathbb{P}(X \geq 2) = \mathbb{P}\left(\frac{X-1}{\sqrt{4}} \geq \frac{2-1}{\sqrt{4}}\right) = \mathbb{P}\left(\frac{X-1}{2} \geq \frac{1}{2}\right)$$

$$= 1 - \mathbb{P}\left(\frac{X-1}{2} \leq \frac{1}{2}\right) = \boxed{1 - \Phi\left(\frac{1}{2}\right)}$$

Fact

$$\Phi(-z) = 1 - \Phi(z)$$

Normal Lookup Tables

- You might have noticed that $\Phi(x)$ doesn't have a closed-form expression. This is why we need to use either computer softwares or lookup tables to obtain values of Φ.
- Here is how we can use a lookup table. Suppose We want to find $\Phi(0.34)$

$$\underbrace{0.3}\overbrace{4}$$

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.3023

Normal Lookup Tables

- Most tables only give values for $z > 0$. How would I find, say, $\Phi(-1)$? Use the property $\Phi(-z) = 1 - \Phi(z)$!
- As practice, find $\Phi(1.24)$, $\Phi(3.0)$, and $\Phi(-2.33)$.

Percentiles

Definition: Percentile

The p^{th} **percentile** (sometimes called the p^{th} **quantile**) of a distribution is defined to be the value π_p such that $\mathbb{P}(X \leq \pi_p) = p$, where X is a random variable that follows the distribution in question.

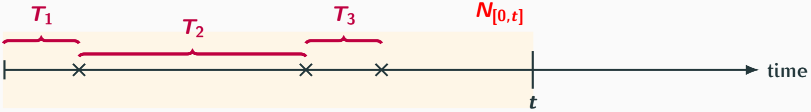
- What other name do we give to the 50^{th} percentile? The Median.

- So, to find the p^{th} percentile we solve the equation $F_X(\pi_p) = p$.
- This is why the inverse of the c.d.f. is sometimes called the **quantile function**.
- Example: quantile function of the Exponential distribution.
- Example: 75^{th} percentile of the standard normal distribution.

Poisson Process, Revisited

Poisson Process, Revisited

- Alright, let's return to our Poisson Point Process again:



- We are finally in a position to find the distribution of the interarrival times. (Spoiler: it will turn out to be a distribution we already know!)
- In words, the event $\{T_1 \geq t\}$ means "the first arrival occurred after time t." Equivalently, what does this say about the number of arrivals in the interval $[0, t]$? There were zero!
- So, what we see is

$$\mathbb{P}(T_1 \geq t) = \mathbb{P}(N_{[0,t]} = 0)$$

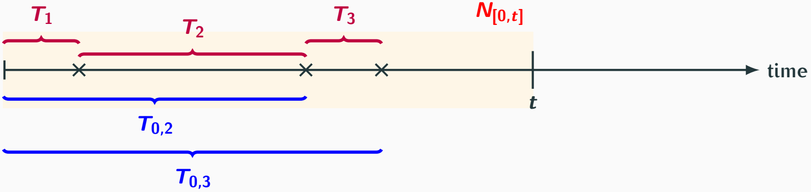
- We know the distribution of $N_{[0,t]}$; it is $\text{Pois}(\lambda \cdot t)$.
- Therefore,

$$\mathbb{P}(T_1 \geq t) = e^{-\lambda t} \cdot \frac{(\lambda t)^0}{0!} = e^{-\lambda t} \implies F_{T_1}(t) = 1 - e^{-\lambda t}$$

Poisson Process, Revisited

- Yup, that's right: $\boxed{T_1 \sim \text{Exp}(\lambda)}$
- In fact, with a bit of work, one can show that $\boxed{T_i \sim \text{Exp}(\lambda)}$ for all i, and that the T_i's are independent. (Loosely speaking, this relates to the memorylessness property along with the fact that the number of arrivals in nonoverlapping intervals were assumed to be independent random variables)

- Let's take this even further.



- These new times are called the **arrival times**; in other words, $T_{0,2}$ denotes the "time until the 2nd arrival"

- Let's derive the distribution of $T_{0,2}$.
- Again, we examine $\mathbb{P}(T_{0,2} \geq t)$; the event $\{T_{0,2} \geq t\}$ means the second arrival occurred at a time later than t meaning $N_{[0,t]} \leq 1$. Therefore

$$\mathbb{P}(T_{0,2} \geq t) = e^{-\lambda t} + (\lambda t)e^{-\lambda t}$$

- Equivalently,

$$F_{T_{0,2}}(t) = 1 - e^{-\lambda t} - (\lambda t)e^{-\lambda t}$$

  and so, differentiating w.r.t. t, we find

$$\begin{aligned} f_{T_{0,2}}(t) &= \lambda e^{-\lambda t} - \left[\lambda e^{-\lambda t} - \lambda^2 t e^{-\lambda t}\right] \\ &= \lambda^2 t e^{-\lambda t} \\ &= \frac{\lambda^2}{\Gamma(2)} t^{2-1} e^{-\lambda t} \end{aligned}$$

- That is, $T_{0,2} \sim \text{Gamma}(2, \lambda)$!
- It turns out that $T_{0,n} \sim \text{Gamma}(n, \lambda)$; in other words, the time of arrival of the n^{th}arrival is distributed as a $\text{Gamma}(n, \lambda)$ distribution.
- Additionally, it also turns out that the distribution of the time between the n^{th} and $(n+k)^{\text{th}}$ arrivals is $\text{Gamma}(k, \lambda)$.

Poisson Process, Revisited

- So, here are some summarizing facts. If arrivals follow a Poisson Point Process with rate λ, then:
  - The number of arrivals in a time interval of length t is distributed according to a Pois(λt) distribution.
  - The interarrival times are distributed as Exp(λ) (i.e. the distribution of the times between consecutive arrivals)
  - The arrival times follow the Gamma distribution; specifically, the distribution of the time between the n^{th} and $(n+k)^{\text{th}}$ arrivals is Gamma(k, λ).

- You will discuss Poisson Processes in much greater detail in PSTAT 160B.

Example

Suppose calls arrive at a call center according to a Poisson Process with an average rate of 2 calls per minute.

(a) What is the probability of observing exactly 2 calls between 1pm and 1:01pm? (**Already answered**)

(b) What is the expected number of calls arriving between 2pm and 2:10pm? (**Already answered**)

(c) What is the distribution of the time (in minutes) until the 1$^{\text{st}}$ call?

(d) On average, what is the length of time (in minutes) between the 3$^{\text{rd}}$ and 5$^{\text{th}}$ calls?

Part (c): Exp(2)

Part (d): Let T denote the time between the 3$^{\text{rd}}$ and 5$^{\text{th}}$ calls; then $T \sim \text{Gamma}(2, \lambda)$

$$\mathbb{E}[T] = \frac{2}{2} = \text{1 minute}$$