Week 5

Statistics 251

Zhongjian Wang

Department of Statistics
The University of Chicago

Where are we?

Continuous random variables

We say that X is a **continuous** random variable if there is a **probability density function** $f(x) = f_X(x)$ on $\mathbb{R}$ such that for any $A \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} \mathbf{1}_A(x) f_X(x)dx.$$

Notice that

- $1 = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(x)dx$
- $f_X(x)$ takes non-negative values

Can we have $f_X(x) > 1$ for some x?

Continuous random variables

We say that X is a **continuous** random variable if there is a **probability density function** $f(x) = f_X(x)$ on $\mathbb{R}$ such that for any $A \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} \mathbf{1}_A(x) f_X(x)dx.$$

Notice that

- $\mathbb{P}(X = a) = \int_{\{a\}} f_X(x)dx = 0$ so the probability of any single point is 0.
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$.

The **cumulative distribution function** is

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)dx.$$

So, given F_X is differentiable at x_0,

$$f_X(x_0) = F_X'(X_0)$$

Example

Suppose that the density function is

$$f_X(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$$

What are the values of:

- $\mathbb{P}(X < 3/2)$
- $\mathbb{P}(X \leq 3/2)$
- $\mathbb{P}(1/2 < X < 3/2)$

What is the CDF of X?

In this case, we say that X is **uniformly distributed** on $[0,2]$.

Where are we?

Continuous random variables

Expectation and variance of continuous random variables

Gaussian random variables

Special case of central limit theorem

Exponential distribution

Gamma Distribution

Cauchy Distribution

Expectation

Recall that if X is discrete, we have that

$$\mathbb{E}[X] = \sum_{x:\mathbb{P}(X=x)>0} x \cdot \mathbb{P}(X = x).$$

To modify it to the continuous setting, we write

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

and correspondingly

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

Variance

If X is a continuous random variable, we define

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 p_X(x) dx.$$

As before, we find that

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

Examples

If the density function is

$$f_X(x) = \begin{cases} 1/2 & x \in [0, 2] \\ 0 & x \notin [0, 2], \end{cases}$$

what is $\mathrm{Var}(X)$?

Where are we?

Tossing coins

How many heads will we get from a million fair coin tosses?

- To first order, the expectation is $500,000$.
- How close will the answer be to the expectation?

Tossing coins

What is the probability of getting k heads from n coin tosses?

$$\mathbb{P}(X = k) = \frac{\binom{n}{k}}{2^n}.$$

Let's see a rough plot... ('l9-video.mp4')

Standard Gaussian random variables

We define a **standard Gaussian** random variable to have density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

This is also called a **normal** random variable.

Notice that

$$\begin{aligned}\left[\int_{-\infty}^{\infty} f_x(x)dx\right]^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}dxdy \\ &= \frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2} r\, d\theta dr \\ &= [-e^{-r^2/2}]_0^{\infty} = 1.\end{aligned}$$

Standard Gaussian random variables

We define a **standard Gaussian** random variable to have density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

By symmetry we compute

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0.$$

We may also compute

$$\mathrm{Var}(X) = \mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1.$$

Gaussian random variables

Let X be standard Gaussian. Consider $Y = \sigma X + \mu$. It has density

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and mean and variance

$$\mathbb{E}[Y] = \mu \qquad \text{and} \qquad \mathrm{Var}(Y) = \sigma^2.$$

Cumulative distribution function

Let X be standard Gaussian. The CDF is

$$F_X(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2} dx.$$

It **does not have** explicit representation. Instead

$$\Phi(x) = F_X(x)$$

is called the **error function**. Some values are:

$$\Phi(-1) \approx 0.159 \qquad \Phi(-2) \approx 0.023 \qquad \Phi(-3) \approx 0.0013.$$

This means ≈ 95 percent of the time a Gaussian is within 1 standard deviation of the mean.

Where are we?

DeMoivre-Laplace Limit Theorem

Let S_n be the number of heads in n tosses of a coin which is heads with probability p. We have

$$\mathrm{Var}(S_n) = np(1-p) \implies \text{std. dev.} = \sqrt{np(1-p)}.$$

The normalized deviations of sample mean away from the expectation is

$$\frac{S_n - np}{\sqrt{np(1-p)}}.$$

Theorem (DeMoivre-Laplace)

We have that

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \Phi(b) - \Phi(a),$$

where $\Phi(b) - \Phi(a) = \mathbb{P}(a \le X \le b)$ *for a standard Gaussian* X*.*

This is why we used normalizing factor in the movie.

Examples

Approximate the probability that we get more than $501,000$ heads in a million fair coin tosses.

- The standard deviation is $\sqrt{np(1-p)} = 500$ and $1000 = 500 \cdot 2$.
- The answer is approximately

$$1 - \Phi(2) = \Phi(-2) \approx 0.023.$$

Do all sets have probabilities? (Optional reading)

Consider the uniform measure on $[0,1]$ with $f_X(x) = \mathbf{1}_{[0,1]}(x)$.

Consider the translations $T_r : [0,1) \to [0,1)$ so that

$$T_r(x) = x + r \pmod 1.$$

Call $x, y \in [0,1)$ equivalent if $x - y$ is rational.
Now we try to find $A \subset [0,1)$ such that each point in $[0,1)$ is equivalent to exactly 1 point in A.

Do all sets have probabilities? (Optional reading)

Consider the uniform measure on $[0,1]$ with $f_X(x) = \mathbf{1}_{[0,1]}(x)$.

Consider the translations $T_r : [0,1) \to [0,1)$ so that

$$T_r(x) = x + r \pmod 1.$$

Call $x, y \in [0,1)$ equivalent if $x - y$ is rational.
Now we try to find $A \subset [0,1)$ such that each point in $[0,1)$ is equivalent to exactly 1 point in A. Such set is called Vitali set.

The existence of Vitali set can be proved by axiom of choice. In modern algebra, $A = \mathbb{R}/\mathbb{Q}$.

Uniform measure on $[0,1]$

For this choice, we have that $T_r(A)$ and $T_s(A)$ are disjoint for $r \neq s$ and that

$$[0,1) = \bigcup_{r\in\mathbb{Q}} T_r(A).$$

Let $A_r = T_r A$, first $\mathbb{P}(A_r) = \mathbb{P}(A)$ and

$$\mathbb{P}([0,1]) = \sum_{r\in\mathbb{Q}} \mathbb{P}(A_r) = \sum_{r\in\mathbb{Q}} \mathbb{P}(A).$$

What does this mean for $\mathbb{P}(A)$?

Not all sets have probabilities

Possible answers:

1. Axioms of mathematics are wrong. In other words, we cannot chose A so that

   each point in $[0, 1)$ is equivalent to exactly 1 point in A.

   This requires something called the **axiom of choice**.
2. Accept that not all sets can have probabilities. Instead, define some sets to be **measurable** and allow only those to have probabilities. In this case, the relevant sets are countable unions of intervals and points.

Most of mainstream mathematics takes Answer 2.

Where are we?

Definition: $X \sim \exp(\lambda)$

Say X is an **exponential random variable of parameter** λ when its probability distribution function is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

For $a > 0$, we have,

$$F_X(a) = \int_0^a f(x)dx = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$$

So that $\mathbb{P}(X < a) = 1 - e^{-\lambda a}$, what is $\mathbb{P}(X > a)$?

Moment Formula

Suppose X is exponential with parameter λ.
What is $\mathbb{E}X^n$? (n is positive integer.)

$$\begin{aligned}\mathbb{E}X^n &= \int_0^\infty x^n \lambda e^{-\lambda x} dx \\ &= -\int_0^\infty n x^{n-1} \lambda \frac{e^{-\lambda x}}{-\lambda} dx + x^n \lambda \frac{e^{-\lambda x}}{-\lambda} |_0^\infty \\ &= \frac{n}{\lambda} \mathbb{E}X^{n-1}\end{aligned}$$

So $\mathbb{E}X = \frac{1}{\lambda}$ $\operatorname{Var} X = \frac{1}{\lambda^2}$.

Minimum of independent exponentials is exponential

If $X_1 \sim \exp(\lambda_1)$, $X_2 \sim \exp(\lambda_2)$, and X_1 and X_2 are independent, then $\min\{X_1, X_2\} \sim \exp(\lambda_1 + \lambda_2)$

Proof:

Consider $\mathbb{P}(X_1 > a) = e^{\lambda_1 a}$ given $a \geq 0$,

Memoryless property

Memoryless property: If X represents the time until an event occurs, then given that we have seen no event up to time a, the conditional distribution of the remaining time till the event is the same as it originally was.
To be more concrete in math words, we need to prove:
If $Y = X - a$, $\mathbb{P}(Y > b|X > a)$ is indepedent of a?

$$
\begin{aligned}
& P(Y > b|X > a) \\
= & \mathbb{P}(X - a > b|X > a) \\
= & \mathbb{P}(X > a + b)/\mathbb{P}(X > a) \\
= & e^{-\lambda b}
\end{aligned}
$$

Example: Roadtrip

Suppose that the number of miles that a car can run before some part wears out is exponentially distributed with an average value of $10,000$ miles.
If the odometer shows the car has already run for $5,000$ miles.
What is the probability that he or she will be able to complete a $5,000$ miles trip without having to replace that part?

Example: Radioactive Decay

Suppose you start at time zero with n radioactive particles. Suppose that each one (independently of the others) will decay at a random time, which follows $\exp(\lambda)$. Let T be amount of time until no particles are left. What are $\mathbb{E}[T]$?

Let T_1 be the amount of time you wait until the first particle decays, T_2 the amount of additional time until the second particle decays, etc., so that $T = T_1 + T_2 + \cdots$.
Then what is distribution of T_i?

So $\mathbb{E}[T] = \lambda^{-1}(1 + 1/2 + 1/3 + \cdots 1/n)$

Where are we?

Definition

A random variable is said to have a gamma distribution with parameters (α, λ), $\lambda > 0$, $\alpha > 0$, if its density function is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\Gamma(\alpha)$, called the gamma function, is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$$

When α is a positive integer, say, $\alpha = n$, the gamma distribution with parameters (α, λ) often arises, in practice as the distribution of the amount of time one has to wait until a total of n events has occurred.

How to calculate $\Gamma(\alpha)$?

Integration by parts!

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty e^{-y} y^{\alpha-1} dy \\ &= -e^{-y} y^{-\alpha-1}|_0^\infty + \int_0^\infty e^{-y}(\alpha-1) y^{\alpha-2} dy \\ &= (\alpha-1)\Gamma(\alpha-1)\end{aligned}$$

So,

$$\begin{aligned}\Gamma(n) &= (n-1)\Gamma(n-1) \\ &= \cdots \\ &= (n-1)(n-2)\cdots 3 \cdot 2\Gamma(1)\end{aligned}$$

Note $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$,
So $\Gamma(n) = (n-1)!$

Propositions

Recall,

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

• Exponential distribution with parameter λ is a gamma distribution with parameter $(1, \lambda)$.

• If X follows a gamma distribution with parameter (α, λ), then $\mathbb{E}[X] = \frac{\alpha}{\lambda}$. $\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$

Where are we?

Cauchy Distribution

A random variable is said to have a Cauchy distribution with parameter θ, $-\infty < \theta < \infty$, if its density is given by

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}$$

If X follows a Cauchy distribution with parameter 0, $\mathbb{E}X$ is undefined!

Example

Suppose that a narrow-beam flashlight is spun around its center, which is located a unit distance from the x-axis. If the flashlight stops randomly at some point and the angle θ with y axis follows uniform distribution, then the position X follows Cauchy distribution.
Why?

$$\begin{aligned} F(x) &= P\{X \leq x\} \\ &= P\{\tan\theta \leq x\} \\ &= P\left\{\theta \leq \tan^{-1} x\right\} \\ &= \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x \end{aligned}$$

Hence, the density function of X is given by

$$f(x) = \frac{d}{dx}F(x) = \frac{1}{\pi\left(1 + x^2\right)} \quad -\infty < x < \infty$$