STAT 2593

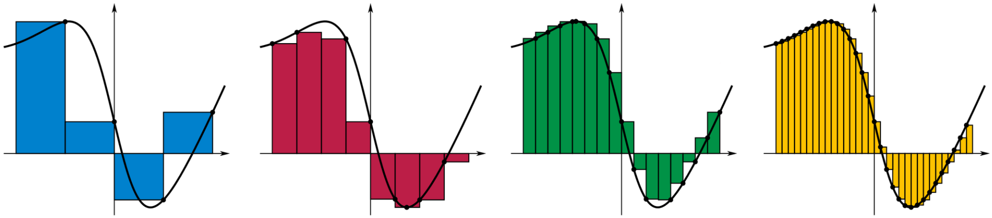
Lecture 017 - Cumulative Distribution Functions and Expected Values

Dylan Spicker

Cumulative Distribution Functions and Expected Values

Learning Objectives

1. Understand how the CDF and PDF are related.
2. Understand how the concepts from discrete random variables (expectation, variance, etc.) translate to continuous random variables.
3. Understand how percentiles relate to continuous distributions.



Discrete versus Continuous Distributions

- Generally, if we replace the summation from the discrete setting with an equivalent integral, we can derive the same quantities.

Discrete versus Continuous Distributions

- Generally, if we replace the summation from the discrete setting with an equivalent integral, we can derive the same quantities.
- For instance, if we want $E[X]$, we take

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

Discrete versus Continuous Distributions

- Generally, if we replace the summation from the discrete setting with an equivalent integral, we can derive the same quantities.
- For instance, if we want $E[X]$, we take

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

  - We get $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$ and $\mathrm{var}(X) = \int_{-\infty}^{\infty}(x - E[X])^2 f(x)dx$.

Discrete versus Continuous Distributions

- Generally, if we replace the summation from the discrete setting with an equivalent integral, we can derive the same quantities.
- For instance, if we want $E[X]$, we take
$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$
  - We get $E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$ and $\text{var}(X) = \int_{-\infty}^{\infty}(x - E[X])^2 f(x)dx$.
- We saw in deriving the PDF that the CDF is also given through integration as
$$F_X(x) = \int_{-\infty}^{x} f(t)dt.$$

Computing Probabilities

- As previously discussed, $P(X = x) = 0$ for continuous random variables.

Computing Probabilities

- As previously discussed, $P(X = x) = 0$ for continuous random variables.
- We can write that

$$P(X \geq a) = P(X > a) = \int_a^\infty f(x)dx = 1 - F(a).$$

Computing Probabilities

- As previously discussed, $P(X = x) = 0$ for continuous random variables.

- We can write that

$$P(X \geq a) = P(X > a) = \int_a^{\infty} f(x)dx = 1 - F(a).$$

  - In general, we **do not** need to be careful about inequalities for continuous random variables.

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.
- With a continuous distribution, we can find the population analogue.

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.
- With a continuous distribution, we can find the population analogue.
- Define $\eta(p)$ to be the $100p$-th percentile.

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.
- With a continuous distribution, we can find the population analogue.
- Define $\eta(p)$ to be the $100p$-th percentile.
  - Then solving $p = F_X(\eta(p))$ gives the value.

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.
- With a continuous distribution, we can find the population analogue.
- Define $\eta(p)$ to be the $100p$-th percentile.
  - Then solving $p = F_X(\eta(p))$ gives the value.
  - This is equivalent to

$$p = \int_{-\infty}^{\eta(p)} f(t)dt.$$

Percentiles of Distributions

- Recall that a p-th percentile of a dataset was the value which divided the data into $p\%$ below the value and $(1-p)\%$ above the value.
- With a continuous distribution, we can find the population analogue.
- Define $\eta(p)$ to be the $100p$-th percentile.
  - Then solving $p = F_X(\eta(p))$ gives the value.
  - This is equivalent to

$$p = \int_{-\infty}^{\eta(p)} f(t)dt.$$

- Finding $\eta(0.5)$ gives the median (50-th percentile).

Summary

- Generally, summations can be replaced by integrals and have results maintained.
- The expectation, variance, and CDF are all defined analogously to the discrete case.
- Percentiles can be analytically computed for continuous distributions by inverting the CDF.