Lecture 5

David Armstrong

STATS 67 - UCI

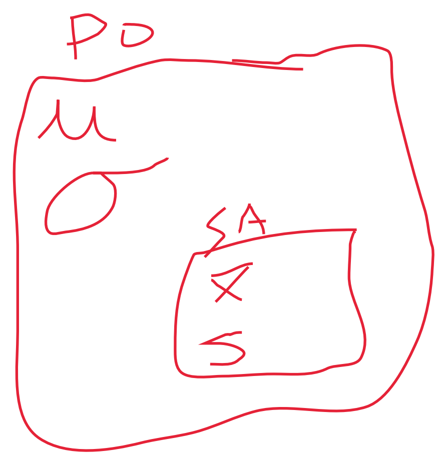
Discrete Random Variables: pdf & pmf

- $f(x)$ is the probability mass function of a random variable X.
- The input is x, which is a specified value of X from its support.
- The output is $P(X = x)$, the probability that X is equal to what we specified, x.
- Thus $\boxed{f(x) = P(X = x)}$.
- Since it is a probability, $0 \leq f(x) \leq 1$.
- To be a valid probability mass function (pmf), The sum of the pmf's must sum to one.
  - $\sum_{x \in \mathbb{S}_X} f(x) = 1$.
- The cumulative distribution function (cdf) of X is not just $P(X = x)$ but $\boxed{P(X \leq x)}$.
- We denote cdf as F(x) (where the pmf is f(x)).
- $F(x) = P(X \leq x) = \sum_{\tilde{x} \leq x} f(\tilde{x}) = \sum_{\tilde{x} \leq x} P(X = \tilde{x})$.
- The cumulative distribution function is the sum of several probability mass functions.
- Note that $F(x) = P(X \leq x) = 1 - P(X > x)$.

P

Discrete Random Variables: Expectation

- The *expectation* of X is denoted as $\boxed{\mathrm{E}(X)}$.
- $\mathrm{E}(X) = \sum_{x \in \mathbb{S}_X} xP(X = x) = \sum_{x \in \mathbb{S}_X} xf(x)$.
- The expectation be viewed as averaging over all possible X values while weighting each possible value by its probability.
- For ease of notation, we let μ be the population expected value of X. That is to say $\mathrm{E}(X) = \mu$.



Discrete Random Variables: Variance

Now we come to another quantity that describes the probability distribution of X, known as the *variance*.

- The *variance* of X is defined to be the average squared deviation from the mean.
- $\text{Var}(X)=\text{E}[(X-\text{E}(X))^2]=\text{E}[(X-\mu)^2]$.
- $\boxed{\text{Var}(X)= \text{E}(X^2)-[\text{E}(X)]^2}$
- We denote the variance of X as σ^2.
- Since it is the expected value of a squared random variable, $\sigma^2 > 0$.
- $\boxed{\sigma = \sqrt{\text{Var}(X)}}$ is known as the *standard deviation* of X.
  - σ - The typical distance of the datapoints to the mean.

Properties

$E(c) = c$

- If c is a constant, then $\text{Var}(c)= 0$
- If c is a constant, then $\text{Var}(cX) = c^2\text{Var}(X)$

$E(cX) = cE(X)$

If X and Y are independent random variables.

- $\text{Var}(X+Y)=\text{Var}(X)+\text{Var}(Y)$

As a result, if X_i's are independent random variables and c_i's are constants, then:

- $\text{Var}(\sum_i c_i X_i) = \sum_i c_i^2 \text{Var}(X_i)$

Discrete Random Variables: Variance

- Example:
  Let $\mathbb{S}_X = \{-1, 1\}$ where $P(X = 1) = 0.5$ and $P(X = -1) = 0.5$
  - Calculate the expectation of X.

    $E(X) = \sum x f(x)$

  - Calculate the variance of X

- Example:
  Let $\mathbb{S}_X = \{-1000, 1000\}$ where $P(X = 1000) = 0.5$ and $P(X = -1000) = 0.5$
  - Calculate the expectation of X.
  - Calculate the variance of X

Discrete Random Variables: Variance

Assume we roll a fair die once. Let X is equal to the number that shows on the die. Below is a relative frequency table.

	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$	$x = 6$
$f(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- What is the variance of X?

- What is the standard deviation of X?

Discrete Random Variables: Distribution

Assume we flip a fair coin 3 times. A distribution table for X, the number of heads, is posted below.

	$x=0$	$x=1$	$x=2$	$x=3$
$f(X=x)$	$\left(\frac{1}{2}\right)^3$	$3\left(\frac{1}{2}\right)^3$	$3\left(\frac{1}{2}\right)^3$	$\left(\frac{1}{2}\right)^3$

- What is the expected value of X?

- Calculate the Variance of X

- Calculate $\mathrm{Var}(5+4X)$

Discrete Random Variables

- Statisticians have been studying distributions for a long time.
- We are trying to find the probability of seeing a specific observation from a sample, given some population parameters.
- We say $X \sim Distribution(\Phi)$
- We define Φ as representing the population parameters.

Discrete Random Variables

Example

- Set $X = \{0, 1\}$.
- Let $P(X = 1) = p$ and $P(X = 0) = 1 - p$.
- Calculate the expectation of X.

- Calculate the variance of X.

Bernoulli Distribution

- Assume we perform one event.
- If we set $X = \{0, 1\}$.
- Let p be the probability of success.
- Let $P(X = 1) = p$ and $P(X = 0) = 1 - p$.
- Then we say X follows a Bernoulli distribution with parameter p
- Denoted $X \sim Bernoulli(p)$
- $f(x) = P(X = x) = p^x(1 - p)^{1-x}$
  - $\mathrm{E}(X) = p$.
  - $\mathrm{Var}(X) = p(1 - p)$.

Example: Assume we perform an experiment with one trial where probability of success is 42%. Write out a relative frequency table for the event.

Discrete Random Variables

- **Example:** Assume we perform an experiment with 4 repeated trials where the probability of success for each trial is p.
- What is the support of X?
- How many total outcomes ar there?
- Draw out a table of all possible outcomes for each element in the support.
- Below each column, writh probability of each element in the **Support**

Binomial Distribution

- The binomial random variables with n trials and p parameter can be characterized as the number of success' in n independent trials.
- We define the probability of success on any given trial to be p, and the probability of no success on a given trial to be $1-p$.
- The probability of each independent event does not change. p is constant
- Then we say X follows a Binomial distribution with parameters n and p.
- Denoted $X \sim Binomial(n,p)$
  - $\mathrm{E}(X) = np$.
  - $\mathrm{Var}(X) = np(1-p)$.
- $f(x) = P(X=x) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x(1-p)^{n-x} \qquad \text{for } x = 0, 1, 2, ..., n.$

R Code

- The pmf of the binomial is dbinom(x, n, p) = P(X=x).
- The cdf of the binomial is pbinom(x, n, p) = P(X≤x).

Discrete Random Variables

- **Examples of Binomial random variables.**
- The number of heads to show up when flipping a fair coin 1000 times. ($n = 1000$ and $p = 0.5$).
- A company has 123 employees. All employees are independent of one another, and the probability that a single employee has certification is p. Then the number of employees that have certificates (out of the 123) is a binomial random variable.
- You roll a fair die 21650 times. The number of times a 3 shows is a binomial random variable. Here, $p = \frac{1}{6}$ is the probability of seeing a 3 on a given roll. The probability of not seeing a 3 (1,2,4,5, or 6) is $1 - p = \dfrac{5}{6}$.

Discrete Random Variables

Example

Let X be a random variable that follows a binomial distribution with $n = 100$ and $p = \frac{1}{4}$

- What is the distribution of X
- Calculate the expectation of X.
- Calculate the variance of X.
- Calculate the standard deviation of X.

Discrete Random Variables

Example

Let X be a random variable that follows a binomial distribution with $n = 100$ and $p = \frac{1}{4}$

- What is the probability of seeing 5 successes?
- What is the probability of seeing 55 successes?
- What is the probability that we see less than 5 successes?
- What is the probability that we see at most 3 successes?
- What is the probability that we see less than 99 successes?

Discrete Random Variables

Example According to CTIA, 32% of all U.S. housholds are wireless-only households, meaning they have no landline. In a random sample of 20 households, what is the probability that:

- Exactly 5 are wireless-only?
- Fewer thatn 3 are wireless-only?
- How many households in your sample would you expect to be wireless-only?
- What is the standard deviation of homes in your sample that would be wireless-only?