Lecture 6

David Armstrong

STATS 67 - UCI

Discrete Random Variables

- The discussion up to now was about bounded discrete random variables.
- We will extend to the case of unbounded discrete random variables.
- An unbounded discrete random variable has uncountably many possible outcomes.
- Example is $\mathbb{S}_X = \{0, 1, 2, 3, ...\}$ or $\mathbb{S}_X = \{1, 2, 3, ...\}$.
- Examples:
  - The number of cars crossing a bridge in a given time period.
  - The number of white blood cells created in a patient while sick.
  - The number of attempts needed before a success is seen.
- In all examples, there is no obvious upper bound to X,
  thus $X = \{1, 2, ..., \infty\}$.

Discrete Random Variables

Geometric Distribution Assumptions

- Let X be the random variable which is the number of attempts needed before the first success is observed.
- The attempts are assumed to be independent, with a constant probability of success p.
- This is similar to the binomial distribution, but now X is the number of times we attempt before we get the first success.
- The Bernoulli trials are independent, with probability p of success (of a 1).
  - $P(X = 1) = p$, you get a success on the first try.
  - $P(X = 2) = (1 - p)p$, you get a fail on the first try, and success on 2nd.
  - $P(X = 3) = (1 - p)^2 p$, you get a fail on the first two tries, and success on 3rd.
- $\mathbb{S}_X = \{1, 2, 3, ...\}$.
- The probability distribution of the number X of Bernoulli trials needed to get one success.
- This distribution is the *geometric distribution.*

Geometric Distribution

- Let X be the random variable which is the number of attempts needed before the first success is observed. The first success occurs for the value of x.
- The attempts are assumed to be independent, with a constant probability of success p.
- Denoted $X \sim Geometric(p)$
- $f(x) = P(X = x) = (1-p)^{x-1}p$.
  - This to say that we have $x - 1$ fails and the x-th trial is a success.
- F$(X) = P(X \leq x) = 1 - (1-p)^x$
- E(X)=$\frac{1}{p}$
- Var(X)=$\frac{1-p}{p^2}$

R-Code

- $P(X = x) = dgeom(x-1, p)$
- $P(X \leq x) = pgeom(x-1, p)$

Discrete Random Variables

- Some key features of the **geometric series**. Assume $|r| < 1$.
  - $g(r) = \sum_{k=0}^{\infty} ar^k = a(1-r)^{-1} = \dfrac{a}{1-r}$.
  - $\dfrac{d}{dr} g(r) = \sum_{k=1}^{\infty} akr^{k-1} = a(1-r)^{-2} = \dfrac{a}{(1-r)^2}$.
  - $\dfrac{d^2}{dr^2} g(r) = \sum_{k=2}^{\infty} ak(k-1)r^{k-2} = 2a(1-r)^{-3} = \dfrac{2a}{(1-r)^3}$.
  - $g(r) = \sum_{k=0}^{n-1} ar^k = \dfrac{1-r^n}{1-r}$

Proof for the expectation of X.

- Let X follow a geometric distribution with parameter p.

$$
\begin{aligned}
E(X) &= \sum_{x \in \mathbb{S}_X} x f(x) \\
&= \sum_{x=0}^{\infty} x(1-p)^{x-1} p \\
&= p \sum_{x=0}^{\infty} x(1-p)^{x-1} \\
&= p \frac{1}{[(1-(1-p)]^2} \\
&= p \left(\frac{1}{p^2} \right) \\
&= \frac{1}{p}
\end{aligned}
$$

Discrete Random Variables

To prove the variance of X, we need a couple of tricks.

- Within one side of an expression, we are able to add and subtract a constant.
- We will add and subtract a term of E(X) to the right side of the variance formula.
- From our understanding of geometric series:

$$-\sum_{x=2}^{\infty} ax(x-1)r^{x-2} = 2a(1-r)^{-3} = \frac{2a}{(1-r)^3}.$$

- The Proof of the variance for a geometric random variable

$$\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= E(X^2) - E(X) + E(X) - [E(X)]^2 \\
&= E(X^2 - X) + E(X) - [E(X)]^2 \\
&= E[X(X-1)] + E(X) - [E(X)]^2 \\
&= \sum_{x=2}^{\infty} x(x-1)(1-p)^{x-1}p + E(X) - [E(X)]^2 \\
&= \frac{2(1-p)}{p^2} + E(X) - [E(X)]^2 \\
&= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\
Var(X) &= \frac{1-p}{p^2}
\end{aligned}$$

Discrete Random Variables

Example: A kidney patient is waiting for a suitable donor match. Assume you are sampling a big dataset, and the probability that any given person is match for this patient is 0.02. Assume the samples are independent.

- What is the value of p? Assume there is a random variable X, describe its meaning in words. What is the distribution of X

- What is the probability you will have to look at 10 donors before you get to the first match?

- What is the probability it will take less than 15 donors before a match is made?

- What is the expected number of trials needed before the first match is made?

- What is the variance of the number of trials needed before the first match?

Discrete Random Variables

To introduce another unbounded discrete random variable distribution.

- We now come to the *Poisson distribution*.
- The random variable X that follows a Poisson distribution can be seen as the number of events in an interval of time.
- The support of X is $\mathbb{S}_X = \{0, 1, 2, 3, ...\}$.
- We can view the time interval being subdivided into much smaller evenly spaced intervals.
- Example:
  - The number of cars crossing a bridge in a given day (24 hour period).
  - Think of dividing the 24 hours into 1 minute intervals.
  - This is the 1440 minutes in a day separated into 1 minute intervals.

Poisson distribution

- The Poisson random variable is the number of events seen across these many time intervals.
- The time subintervals are assumed to be independent.
- Furthermore, the probability of a success in each time subinterval is constant.
- Let λ denote the average number of events per time interval.
- Denoted $X \sim Poisson(\lambda)$
- $f(x) = P(X = x) = \dfrac{e^{-\lambda}\lambda^{x}}{x!}$.
- $\mathrm{E}(X) = \lambda$
- $\mathrm{var}(X) = \lambda$
- $F(x) = P(X \leq x) = \sum_{\tilde{x}=0}^{x} \dfrac{e^{-\lambda}\lambda^{\tilde{x}}}{\tilde{x}!}$.

R Code

- The pmf of the Poisson is dpois(x, λ) = P(X=x).
- The cdf of the Poisson is ppois(x, λ) = P(X≤x).

Discrete Random Variables

Poisson distribution.

- To formalize this concept, think of the time interval of interest is being divided into n many equally spaced intervals.
- We can think of this as dividing a day into n many intervals.
- We define the probability of an event in a given sub interval (one of the n many) is $\frac{\lambda}{n}$, where $\lambda > 0$.
- We can think of it as $p_n = \frac{\lambda}{n}$.
- p_n is the probability of an event in a time interval (out of n many).
- $p_n = \frac{\lambda}{n} \to 0$ as $n \to \infty$.
- If the time sub intervals are assumed to be independent, then one way to think of the situation is as a binomial distribution (the number of events in a given time interval).
- $P(X = x) = f(x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$

Tools for Poisson Proof:

- $\left(1 + \frac{y}{n}\right)^n \to e^y$ as $n \to \infty$.
- $e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!}$
- $\frac{n!}{(n-x)!} \frac{1}{n^x} \to 1$ as $n \to \infty$.

Discrete Random Variables

Poisson distribution.

- Assume we let $n \to \infty$ (dividing the time interval into smaller and smaller sub intervals).
- Taking the limit of a product is equal to the product of the limits.

$$P(X = x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$P(X = x) = \frac{n!}{x!(n-x)!}\left(\frac{\lambda^{x}}{n^{x}}\right)\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$P(X = x) = \frac{n!}{(n-x)!}\frac{1}{n^{x}}\left(1-\frac{\lambda}{n}\right)^{n-x}\left(\frac{\lambda^{x}}{x!}\right)$$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \text{ as } n \to \infty \ .$$

- This is the probability mass function of the Poisson distribution.
  - The parameter is $\lambda > 0$.
- $\mathbb{S}_X = \{0, 1, 2, 3, ...\}$.

Discrete Random Variables

Proof for the expectation of X.

$$
\begin{aligned}
E(X) &= \sum_{x \in \mathbb{S}_X} x f(x) \\
&= \sum_{x=0}^{\infty} x \, \frac{e^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} x \, \frac{\lambda^x}{x!} \\
&= e^{-\lambda}\lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= e^{-\lambda} e^{\lambda} \lambda = \lambda
\end{aligned}
$$

The Poisson distribution can be thought of modeling the number of events in a given time interval.

Examples:

- Number of trucks crossing a toll bridge in a given day.
- How many requests are given in a network in an hour.
- The number of users that visit a website in a week.
- How many admissions a hospital department has in a given day.

Discrete Random Variables

Example: The number of phone calls to a call service center follows a Poisson distribution with a expected number of 100 calls an hour.

- What is the value of λ?

- What is the probability that 75 calls will be made in an hour?

- What is the expected value of X?

- What is the variance of X?

- What is the standard deviation of X?

Discrete Random Variables

A useful property of the Poisson distribution is that it is infinitely divisible.

- This is the notion that a random variable can be written as the sum of n many (where n can get infinitely large) independent random variables.
- With the Poisson distribution, a given Poisson random variable with parameter λ can be written as the sum of n many independent Poisson random variables where each one has parameter $\dfrac{\lambda}{n}$.
- This implies that the sum of independent Poisson random variables each with λ_i parameter is a Poisson random variable with parameter $\sum_i \lambda_i$.

Discrete Random Variables

Example: Let X represent the number of cars that cross a given bridge in an hour where the average number of cars is 50 an hour. Assume X is a Poisson random.

- What is the value of λ?

- Let Y be a poisson random variable representing the number of cars that cross the same bridge in 2 hours. What is λ_Y?

- Let Z be a poisson random variable representing the number of cars that cross the same bridge in half an hour. What is λ_Z?