Special Discrete Random Variables (Ch. 5.1)

Will Landau

Iowa State University

Feb 19, 2013

Outline

Binomial Distribution

Geometric Distribution

Poisson Distribution

The Binomial Distribution

- $X \sim \text{Binomial}(n,\ p)$ − i.e., X is distributed as a binomial random variable with parameters n and p $(0 < p < 1)$ if:

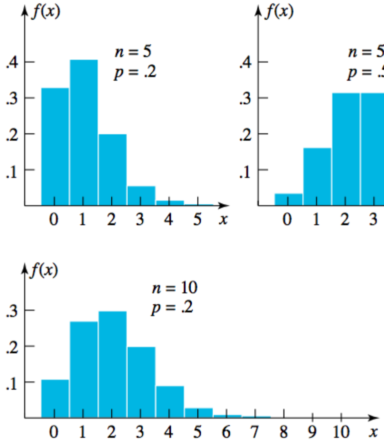
$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

  where:

  - $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, read "$n$ choose x"
  - $n! = n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$, the factorial function.

- $E(X) = np$
- $Var(X) = np(1-p)$

The Binomial Distribution



Purpose of the binomial random variable

- A Bin(n, p) random variable counts the number of successes in n success-failure trials that:
  - are independent of one another.
  - each succeed with probability p.
- Examples:
  - Number of conforming hexamine pellets in a batch of $n = 50$ total pellets made from a pelletizing machine.
  - Number of runs of the same chemical process with percent yield above 80%, given that you run the process a total of $n = 1000$ times.
  - Number of rivets that fail in a boiler of $n = 25$ rivets within 3 years of operation. (Note; "success" doesn't always have to be good.)

Example: machine with 10 components

- Suppose you have a machine with 10 **independent** components in series. The machine only works if all the components work.
- Each component succeeds with probability $p = 0.95$ and fails with probability $1 - p = 0.05$.
- Let Y be the number of components that succeed in a given run of the machine. Then:

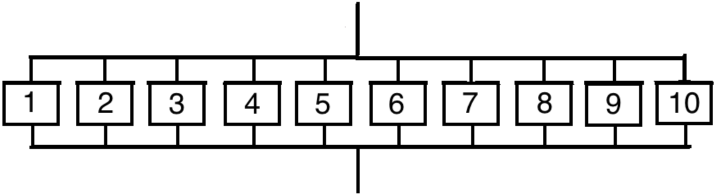
$$Y \sim \text{Binomial}(n = 10, p = 0.95)$$

Example: machine with 10 components

$$
\begin{aligned}
P(\text{machine succeeds}) &= P(Y = 10) \\
&= \binom{10}{10} p^{10} (1-p)^{10-10} \\
&= p^{10} \\
&= 0.95^{10} \\
&= 0.5987
\end{aligned}
$$

- This machine isn't very reliable.

Example: machine with 10 components



- What if I arrange these 10 components in parallel? This machine succeeds if at least 9 of the components succeed.
- What is the probability that the new machine succeeds?

Example: machine with 10 components

$$
\begin{aligned}
&P(\text{improved machine succeeds}) \\
&\quad = P(Y \geq 9) \\
&\quad = P(Y = 9) + P(Y = 10) \\
&\quad = \binom{10}{9} p^9 (1-p) + \binom{10}{10} p^{10} (1-p)^{10-10} \\
&\quad = (10) \cdot 0.95^9 \cdot 0.05 + (1) \cdot 0.95^{10} \\
&\quad = 0.9139
\end{aligned}
$$

- By allowing just one component to fail, we made this machine far more reliable.

Example: machine with 10 components

- If we allow up to 2 components to fail:

$$
\begin{aligned}
&P(\text{improved machine succeeds}) \\
&\quad = P(Y \geq 8) \\
&\quad = P(Y = 8) + P(Y = 9) + P(Y = 10) \\
&\quad = \binom{10}{8} p^8 (1-p)^{10-8} + \binom{10}{9} p^9 (1-p) + \binom{10}{10} p^{10} (1-p)^{10-10} \\
&\quad = \frac{10!}{(10-8)!8!} \cdot 0.95^8 \cdot 0.05^2 + (10) \cdot 0.95^9 \cdot 0.05 + (1) \cdot 0.95^{10} \\
&\quad = 0.9885
\end{aligned}
$$

Example: machine with 10 components

- $E(Y) = np = 10 \cdot 0.95 = 9.5$. So the number of components to fail per run on average is 9.5.
- $Var(Y) = np(1-p) = 10 \cdot 0.95 \cdot (1 - 0.95) = 0.475$.
- $SD(Y) = \sqrt{Var(Y)} = \sqrt{np(1-p)} = 0.689$.

Outline

Geometric random variables

- $X \sim \text{Geometric}(p)$ – that is, X has a geometric distribution with parameter p $(0 < p < 1)$ – if its pmf is:

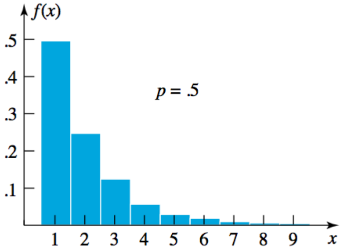
$$f_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

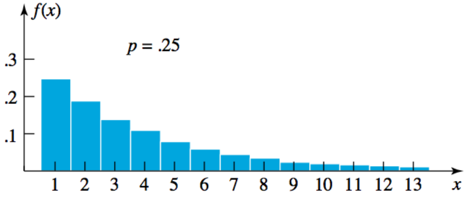
and its cdf is:

$$F_X(x) = \begin{cases} 1-(1-p)^{x} & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- $E(X) = \frac{1}{p}$
- $Var(X) = \frac{1-p}{p^2}$

A look at the Geom(p) distribution





Uses of the $X \sim \text{Geom}(p)$

- For an indefinitely-long sequence of independent, success-failure trials, each with P(success) $= p$, X is the number of trials it takes to get a success.
- Examples:
  - Number of rolls of a fair die until you land a 5.
  - Number of shipments of raw material you get until you get a defective one.
  - The number of enemy aircraft that fly close before one flies into friendly airspace.
  - Number hexamine pellets you make before you make one that does not conform.
  - Number of buses that come before yours.

Example: shorts in NiCad batteries

- An experimental program was successful in reducing the percentage of manufactured NiCad cells with internal shorts to around 1%.
- Let T be the test number at which the first short is discovered. Then, $T \sim \text{Geom}(p)$.

$$
\begin{aligned}
P(\text{1st or 2nd cell tested is has the 1st short}) &= P(T = 1 \text{ or } T = 2) \\
&= f(1) + f(2) \\
&= p + p(1-p) \\
&= 0.01 + 0.01(1 - 0.01) \\
&= 0.02
\end{aligned}
$$

$$
\begin{aligned}
P(\text{at least 50 cells tested w/o finding a short}) &= P(T > 50) \\
&= 1 - P(T \leq 50) \\
&= 1 - F(50) \\
&= 1 - (1 - (1-p)^x) \\
&= (1-p)^x \\
&= (1 - 0.01)^{50} \\
&= 0.61
\end{aligned}
$$

Example: shorts in NiCad batteries

$$E(T) = \frac{1}{p} = \frac{1}{0.01}$$
$$= 100 \text{ tests for the first short to appear, on avg.}$$

$$SD(T) = \sqrt{Var(T)} = \sqrt{\frac{1-p}{p^2}}$$
$$= \sqrt{\frac{1-0.01}{0.01^2}} = 99.5 \text{ tested batteries}$$

Outline

Binomial Distribution

Geometric Distribution

Poisson Distribution

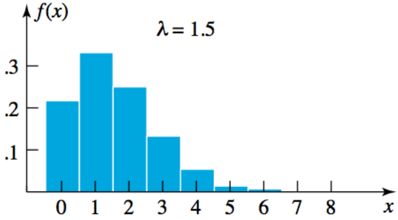
Poisson random variables

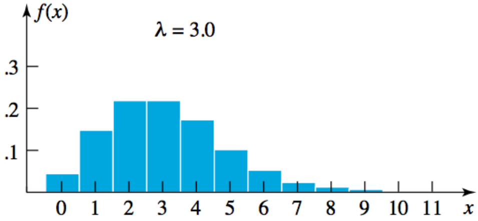
- $X \sim \text{Poisson}(\lambda)$ – that is, X has a geometric distribution with parameter $\lambda > 0$ – if its pmf is:

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- $E(X) = \lambda$
- $Var(X) = \lambda$

A look at the Poisson distribution





Meaning of the Poisson distribution

- A Poisson(λ) random variable counts the number of occurrences that happen over a fixed interval of time or space.
- These occurrences must:
  - be independent
  - be sequential in time (no two occurrences at once)
  - occur at the same constant rate, λ.
- λ, the **rate parameter**, is the expected number of occurrences in the specified interval of time or space.

Examples

- Y is the number of shark attacks off the coast of CA next year. $\lambda = 100$ attacks per year.
- Z is the number of shark attacks off the coast of CA next month. $\lambda = 100/12 = 8.3333$ attacks per month
- N is the number of β particles emitted from a small bar of plutonium, registered by a Geiger counter, in a minute. $\lambda = 459.21$ particles/minute.
- J is the number of particles per three minutes. $\lambda = ?$

$$\begin{aligned}\lambda &= \frac{459.21 \text{ (units particle)}}{1 \text{ (unit minute)}} \cdot \frac{3 \text{ (units minute)}}{1 \text{ (unit of 3 minutes)}} \\ &= \frac{1377.63 \text{ (units particle)}}{1 \text{ (unit of 3 minutes)}} = 1377.62 \text{ particles per 3 minutes}\end{aligned}$$

Example: Rutherford/Geiger experiment

- Rutherford and Geiger measured the number of α particles detected near a small bar of plutonium for 8-minute periods.
- The average number of particles per 8 minutes was $\lambda = 3.87$ particles / 8 min.
- Let $S \sim \text{Poisson}(\lambda)$, the number of particles detected in the next 8 minutes.

$$f(s) = \begin{cases} \frac{e^{-3.87}(3.87)^s}{s!} & s = 0, 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} &P(\text{at least 4 particles recorded}) \\ &\quad = P(S \geq 4) \\ &\quad = f(4) + f(5) + f(6) + \cdots \\ &\quad = 1 - f(0) - f(1) - f(2) - f(3) \\ &\quad = 1 - \frac{e^{-3.87}(3.87)^0}{0!} - \frac{e^{-3.87}(3.87)^1}{1!} \\ &\qquad - \frac{e^{-3.87}(3.87)^2}{2!} - \frac{e^{-3.87}(3.87)^3}{3!} \\ &\quad = 0.54 \end{aligned}$$

Example: arrival at a university library

- Some students' data indicate that between 12:00 and 12:10 P.M. on Monday through Wednesday, an average of around 125 students entered a library at Iowa State University library.
- Let M be the number of students entering the ISU library between 12:00 and 12:01 PM next Tuesday.
- Model $M \sim \text{Poisson}(\lambda)$.
- Having observed 125 students enter between 12:00 and 12:10 PM last Tuesday, we might choose:

$$\lambda = \frac{125 \text{ (units of student)}}{1 \text{ (unit of 10 minutes)}} \cdot \frac{1 \text{ (unit of 10 minutes)}}{10 \text{ (units of minute)}}$$
$$= \frac{12.5 \text{ (units of student)}}{1 \text{ (unit minute)}} = 12.5 \text{ students per minute}$$

Example: arrival at a university library

- Under this model, the probability that between 10 and 15 students arrive at the library between 12:00 and 12:01 PM is:

$$
\begin{aligned}
P(10 \le M \le 15) &= f(10) + f(11) + f(12) + f(13) + f(14) + f(15) \\
&= \frac{e^{-12.5}(12.5)^{10}}{10!} + \frac{e^{-12.5}(12.5)^{11}}{11!} + \frac{e^{-12.5}(12.5)^{12}}{12!} \\
&+ \frac{e^{-12.5}(12.5)^{13}}{13!} + \frac{e^{-12.5}(12.5)^{14}}{14!} + \frac{e^{-12.5}(12.5)^{15}}{15!} \\
&= 0.60
\end{aligned}
$$

Example: shark attacks

- Let X be the number of unprovoked shark attacks that will occur off the coast of Florida next year.
- Model $X \sim \text{Poisson}(\lambda)$.
- From the shark data at http://www.flmnh.ufl.edu/fish/sharks/statistics/FLactivity.htm, 246 unprovoked shark attacks occurred from 2000 to 2009.
- Hence, I calculate:

$$\lambda = \frac{246 \text{ (units attack)}}{1 \text{ (unit of 10 years)}} \cdot \frac{1 \text{ (unit of 10 years)}}{10 \text{ (units year)}}$$
$$= \frac{24.6 \text{ (units attack)}}{1 \text{(unit year)}} = 24.6 \text{ attacks per year}$$

Example: shark attacks

$$P(\text{no attacks next year}) = f(0) = e^{-24.6} \cdot \frac{24.6^0}{0!}$$
$$\approx 2.07 \times 10^{-11}$$

$$P(\text{at least 5 attacks}) = 1 - P(\text{at most 4 attacks})$$
$$= 1 - F(4)$$
$$= 1 - f(0) - f(1) - f(2) - f(3) - f(4)$$
$$= 1 - e^{-24.6}\frac{24.6^0}{0!} - e^{-24.6}\frac{24.6^1}{1!} - e^{-24.6}\frac{24.6^2}{2!}$$
$$- e^{-24.6}\frac{24.6^3}{3!} - e^{-24.6}\frac{24.6^4}{4!}$$
$$\approx 0.9999996$$

$$P(\text{more than 30 attacks}) = 1 - P(\text{at least 30 attacks})$$
$$= 1 - e^{-24.6}\sum_{i=0}^{30}\frac{24.6^x}{x!} = 1 - e^{-24.6} \cdot 4.251 \times 10^{10}$$
$$\approx 0.1193$$

Example: shark attacks

- Now, let Y be the total number of shark attacks in Florida during the next 4 months.
- Let $Y \sim \text{Poisson}(\theta)$, where θ is the true shark attack rate per 4 months:

$$\theta = \frac{24.6 \text{ (units attack)}}{1 \text{ (unit year)}} \cdot \frac{1/3 \text{ (unit year)}}{1 \text{ (unit of 4 months)}}$$
$$= \frac{8.2 \text{ (units attack)}}{1 \text{ (unit of 4 months)}} = 8.2 \text{ attacks per 4 months}$$

Example: shark attacks

$$
\begin{aligned}
P(\text{no attacks next year}) &= f(0) = e^{-8.2} \cdot \frac{8.2^0}{0!} \\
&\approx 0.000275 \\
P(\text{at least 5 attacks}) &= 1 - P(\text{at most 4 attacks}) \\
&= 1 - F(4) \\
&= 1 - f(0) - f(1) - f(2) - f(3) - f(4) \\
&= 1 - e^{-8.2}\frac{8.2^0}{0!} - e^{-8.2}\frac{8.2^1}{1!} - e^{-8.2}\frac{8.2^2}{2!} \\
&- e^{-8.2}\frac{8.2^3}{3!} - e^{-8.2}\frac{8.2^4}{4!} \\
&\approx 0.9113 \\
P(\text{more than 30 attacks}) &= 1 - P(\text{at least 30 attacks}) \\
&= 1 - e^{-8.2}\sum_{i=0}^{30}\frac{8.2^x}{x!} = 1 - e^{-8.2} \cdot 4.251 \times 10^{10} \\
&\approx 9.53 \times 10^{-10}
\end{aligned}
$$