2.4 The Hypergeometric & Multinomial Distributions

- Bernoulli Binomial Geometric Negative Binomial

trials until S

1st S

r th

$N = 30$ 10 red 20 blue $\frac{10}{30}$

interested in red $N_1 = 10$

Hypergeometric Distribution

e.g. pick 5 marble

Out of a population of size N, suppose we have N_1 successes and N_2 failures.

$n = 5$

(note, $N_1 + N_2 = N$, the probability of a success, $p = N_1 / N$)

Define a random variable X:

the number of successes in a random sample of size n.

dependent independent

If sampling is done without replacement, X follows a **hypergeometric distribution.**

if with replacement binomial

$N=30$ $n=5$ $\frac{\binom{10}{x}\binom{20}{5-x}}{\binom{30}{5}}$ total succ sample size

$N_1=10$

Hypergeometric Distribution

prob exact 2 red?

$X \sim Hypergeom(N, N_1, n)$

(remember: $N = N_1 + N_2$)

$$\boxed{f(x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}},}$$

$$\boxed{x \le n, \; x \le N_1, \;\; n-x \le N_2}$$

- $E[X] = n\frac{N_1}{N}$ $\boxed{n \cdot p}$
- $Var[X] = n\frac{N_1}{N}\frac{N_2}{N}\frac{N-n}{N-1}$

$x=2$ $f(2) = \boxed{\frac{\binom{10}{2}\binom{20}{5-2}}{\binom{30}{5}}} =$

dep

w/o rep

indep

w rep

$\binom{20}{x} p^x (1-p)^{20-x}$

Hypergeometric vs Binomial

If instead, sampling is done one at a time with replacement, X ~ Binomial(n,p) e.g.

$p = .3$

- **Binomial:** A magical beer machine gives the user a stout 30% of the time and an IPA 70% of the time. Let X be the number of stouts you get out of 20 beers.

  $X \sim \text{Binom}(n=20, p=.3)$

- **Hypergeometric:** A nice minibar has 9 stouts and 21 IPAs. Let X be the number of stouts you get if you randomly select 20 beers.

  don't replace

  - What is the pmf of X? $f(x) = \dfrac{\binom{9}{x}\binom{21}{20-x}}{\binom{30}{20}}$, $x \leq 9$

    typo

- Roll a die: independent

- Drawing from a bag: replace -- > independent
- Not replacing → dependent

bi

Multinomial Distribution

Similar to binomial distribution, but for more than 2 groups. E.g.

- Color – Red/Green/Blue
- Your Major – Stats/Math/Engineering/Other

Stat / not Stat → binomial

$\frac{10!}{2!\,2!\,6!}$

$(.1)(.1)(.2)(.2)(.7)(.7)(.7)(.7)(.7)(.7)$

Multinomial Distribution

$X = (X_1, X_2, \ldots X_k) \sim Multinomial(n, p_1, p_2, \ldots, p_k)$

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$

$f(x_1, x_2, \ldots x_k) = \frac{n!}{x_1!x_2!\ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$

indep

- $E[X_i] = np_i$
- $Var[X_i] = np_i(1\text{-}p_i)$

Examples

2.4

1 A deck of cards contains 20 cards: 6 red cards and 14 black cards. 5 cards are drawn randomly *without replacement*.

$X \sim HG(20, 6, 5)$

$N=20, N_1=6, n=5$

What is the probability that exactly 4 red cards are drawn?

$$f(4) = \frac{\binom{6}{4}\binom{14}{1}}{\binom{20}{5}}$$

$X \sim HG(N=20, N_1=14, n=5)$

What is the probability that at least 2 black cards are drawn?

$$P[X \geq 2] = 1 - \underbrace{P[X \leq 1]}_{F(1)} = 1 - \underbrace{f(0) - f(1)}$$

$$= 1 - \frac{\binom{14}{0}\binom{6}{5}}{\binom{20}{5}} - \frac{\binom{14}{1}\binom{6}{4}}{\binom{20}{5}}$$

2 Suppose the majors of students taking Stat 400 can be broken down as follows:

	X_1		X_2		X_3
Math		**Statistics**		**Other**	
10%	p_1	20%	p_2	70%	p_3

$n=10$ Out of 10 randomly sampled students, calculate the probability that this group contains:

- A) $\boxed{2}$ Math, $\boxed{2}$ Stats, and 6 Other $\Rightarrow P[X_1=2, X_2=2, X_3=6]$
  - x_1 x_2 x_3
- B) At least one Stats student $p=.2$

$$\boxed{\frac{10!}{2!\,2!\,6!}\, .1^2\, .2^2\, .7^6}$$

$\rightarrow X \sim Binom(10, .2)$

$P[X \geq 1] = 1 - P[X=0]$

- $E[X_i] = np_i$

3 When Iron Man and Captain America play 

against each other, Iron Man wins 40% of the time, loses 35% of the time and draws 25% of the time. Assume results of games are independent.

$p_3 = .25$

$n=12$

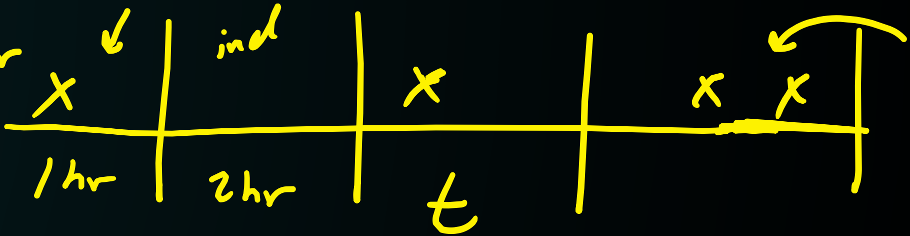
If they play 12 games, what is the probability that Iron Man wins 7, loses 2, and draws 3 games?

$$f(x_1, x_2, \ldots x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1}\, p_2^{x_2} \ldots p_k^{x_k} = \frac{12!}{7!2!3!} \times .40^7 \times .35^2 \times .25^3 = 0.0248$$

If they play 12 games, what is the expected value of the number of games that they will tie? $np_3 = 12(.25) = 3$

2.6 The Poisson Distribution

$\boxed{\lambda = 10}$ / min $\boxed{\lambda = 600}$ /hr



Poisson Process

Definition 2.6-1

Let the number of occurrences of some event in a given continuous interval be counted. Then we have an **approximate Poisson process** with parameter $\lambda > 0$ if the following conditions are satisfied:

(a) The numbers of occurrences in nonoverlapping subintervals are independent.

(b) The probability of exactly one occurrence in a sufficiently short subinterval of length h is approximately λh.

(c) The probability of two or more occurrences in a sufficiently short subinterval is essentially zero.

Poisson Process Examples

- # of cell phone calls passing through a relay tower between 9 and 11 a.m.
- Number of customers that show up to Oberweis between 5-6pm.
- Number of defects in a 100-foot roll of aluminum screen that is 2 feet wide.

Poisson Distribution

$X \sim Poisson(\lambda)$

$X \sim \text{Poisson}(\lambda = 5)$

rate

$\sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = E[X]$

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0,1,2,\ldots$$

$f(3) = \frac{5^3 e^{-5}}{3!} = 0.14$

- $E[X] = \lambda$
- $Var[X] = \lambda$

$P[X \leq 8] = ?$

ppois(8,5) = 0.931

dpois(3,5) = 0.14

Note: λ is the Poisson rate.

Poisson Parameter Scaling

If events occur according to a Poisson process with rate λ, then the rate for a Poisson process in an interval of length t is λt.

$\lambda=3$

e.g. :

minute

Every minute, cars pull up to a drive-through according to a Poisson process with rate $\lambda = 3$.

- In an interval of length 1 hour, the rate is $\lambda = 180$.

3 * 60 (minutes in an hour) = 180

Examples

2.6

in R: d → density
pmf
pdf

p → prob. (cumulative)
cdf

2

Suppose Albert receives spam mail every day according to a Poisson Process with an expected value of 10.

What is the distribution of X? $X \sim Poisson(\lambda = 10)$

$P[X=8]$

$E[X]$

What is the probability that Albert receives 8 items of spam in a given day?

$f(8) = \frac{e^{-10} 10^{8}}{8!} = 0.112599$

What is the probability that Albert receives 10 items of spam in a given day?

$f(10) = \frac{e^{-10} 10^{10}}{10!} = 0.12511$

year: $\lambda = 3650$

2 Suppose Albert receives spam mail every day according to a Poisson Process with an expected value of 10.

$X \sim Pois(10)$

$\boxed{\lambda = 10}$

$\lambda \to \lambda t$

~~Find P[Albert receives 10 items of spam in a given day].~~

hour

$t = \frac{1}{24}$

Find P[0 items of spam in a given day]? $f(0) = \frac{e^{-10}\, 10^{0}}{0!} = e^{-10}$

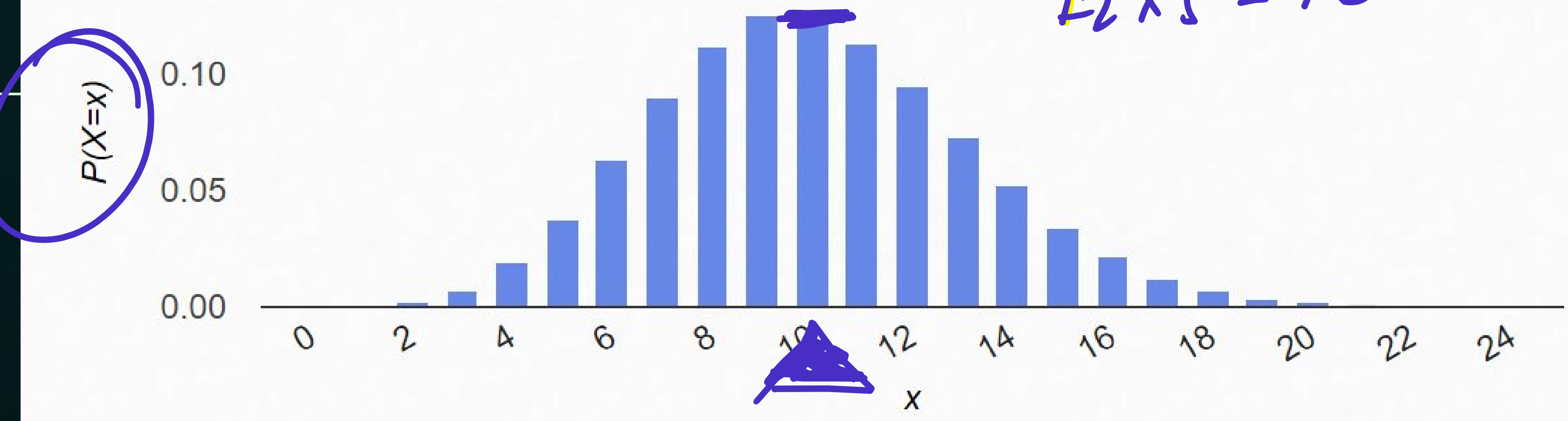
Find P[Albert receives 1 item of spam in a given hour]?

$X \sim Pois\left(\lambda = \frac{10}{24}\right) \leftarrow X \sim Pois(.416\bar{6})$

$f(1) = \frac{e^{-.4166}\, .4166^{1}}{1!} = .2747$

$P[X=10] = ?$

$E[X] = 10$



$$\mu = E(X) = 10 \quad \sigma = SD(X) = 3.162 \quad \sigma^2 = Var(X) = 10$$