Common Discrete Distributions: Bernoulli, Binomial, and Poisson

BIOS 6611

CU Anschutz

Week 2

Discrete Random Variables

Random Variable

- **Random Variable**: variable whose values depend on a random phenomenon. Possible values represent possible outcomes of a yet-to-be-performed experiment.
- We call these possible values the **sample space**
- Notation: X denotes a random variable

Discrete Random Variable

- **Discrete random variable**: sample space is a discrete list of values
  - Ex: Let X denote random experiment of flipping a coin.The possible outcomes are $\{Heads, Tails\}$.
- Characterized by a **probability mass function (PMF)**: function that gives probability that X is equal to values in sample space.
  - Values must be non-negative and sum to 1
  - Notation:

$$
\begin{aligned}
P(X = k) &= \text{function that depends on k,} \\
k &= \{\text{sample space}\}
\end{aligned}
$$

PMF Example

$$
\begin{aligned}
P(X=1) &= 0.2 \\
P(X=3) &= 0.5 \\
P(X=7) &= 0.3
\end{aligned}
$$

$$k = \{1, 3, 7\}$$

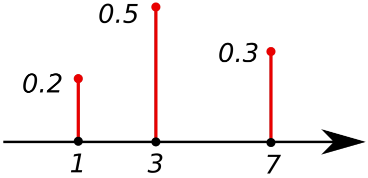


Figure 1: Example PMF (Source: Wiki)

Bernoulli Distribution

Bernoulli Random Variables

- **Bernoulli Random Variable**: sample space is two discrete values
  - Ex: flipping a coin (heads, tails), whether a good pupper catches a treat (success, fail)
- Call one outcome "success", though this can be arbitrary
- Often let 1 represent "success", 0 represent "failure"
- One parameter, p, gives probability of success ($0 \leq p \leq 1$)

Bernoulli PMF:

$$P(X = k) = p^k (1-p)^{1-k}, \quad k = \{0, 1\}$$

$$\begin{aligned} P(X = 1) &= p \\ P(X = 0) &= 1 - p \end{aligned}$$

Binomial Distribution

Binomial Random Variables

- Say want to count number of successes in many Bernoulli trials
- Number of successes is a **Binomial Random Variable**
  - Ex: If I throw 5 treats at my pup, how many do they catch?
- Two parameters: p is probability of success in one trial ($0 \leq p \leq 1$), n is total number of trials ($n \in \{0, 1, 2, \dots\}$)
- Bernoulli is special case of Binomial when $n = 1$

Binomial PMF:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

$$\begin{aligned} P(X = 0) &= (1-p)^n \\ P(X = 1) &= np^1(1-p)^{n-1} \\ P(X = 2) &= \binom{n}{2} p^2 (1-p)^{n-2} \\ &\dots \end{aligned}$$

Why $\binom{n}{k}$?

Q: Why need $\binom{n}{k}$ in Binomial PMF?

A: To account for fact that there are multiple ways to get k successes.

For example, say we throw our pup 3 treats, so $X \sim Binom(3, p)$. Want to know probability that they catch 2 (or 2 "successes"), so want $P(X = 2)$. There are $\binom{n}{k} = \binom{3}{2} = 3$ ways this could happen: catches on first and second, first and third, or second and third throws.

In general, $\binom{n}{k}$ is the number of ways to choose k elements from n elements, in any order.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{1}$$

Binomial Distribution in R

- Let's simulate 10,000 samples from $X \sim Binom(n = 8, p = 0.5)$
- `rbinom` in R.

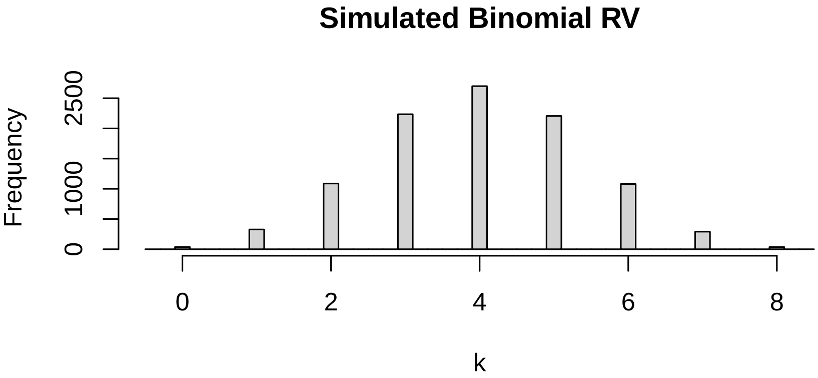
```
# For help documentation, type ?rbinom

# Simulate data
set.seed(812)

# Note here, n is number of simulated samples
# size is number of trials
binom_sample <- rbinom(n = 10000, size = 8, prob = 0.5)
```

Binomial Distribution in R

```
hist(binom_sample, main="Simulated Binomial RV",
     xlab="k", breaks = seq(-0.5, 8.5, 0.2))
```



Other functions for Binomial Distribution in R

- `dbinom`: PMF (d for "density")

$$P(X = k)$$

- `pbinom`: cumulative distribution function, CDF (p for "probability")

$$P(X \leq k) = P(X = 0) + P(X = 1) + \ldots + P(X = k)$$

- `qbinom`: quantile function, smallest value of k such that

$$P(X \leq k) \geq p$$

Other functions for Binomial Examples

- $X \sim Binom(n = 8, p = 0.5)$.
- $P(X = 2) \leftrightarrow$ `dbinom(x = 2, size = 8, prob = 0.5)`
- $P(X \leq 2) \leftrightarrow$ `pbinom(q = 2, size = 8, prob = 0.5)`

```
# PMF
dbinom(x=2, size=8, prob=0.5)
```

```
## [1] 0.109375
```

```
# CDF
pbinom(q=2, size=8, prob=0.5)
```

```
## [1] 0.1445313
```

```
dbinom(x=2, size=8, prob=0.5)+dbinom(x=1, size=8, prob=0.5)+
  dbinom(x=0, size=8, prob=0.5)
```

```
## [1] 0.1445313
```

Poisson Distribution

Poisson Random Variables

- **Poisson Random Variables**: number of "successes" that occur in a given time or space.
- Sample space is discrete and infinite: $k = 0, 1, 2, 3, \ldots$
- Often used for rare events
- Ex: number of meteorites greater than 1 meter diameter that strike Earth each year, infectious cells in well
- Characterized by rate parameter, $\lambda \in (0, \infty)$
- R functions: `dpois`, `ppois`, `qpois`, `rpois`

Poisson PMF:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \ldots \tag{2}$$

Poisson approximation to the binomial

If $X \sim Binom(n, p)$ with n "large" ($n \geq 100$) and p "small" ($p \leq 0.01$), then X can be approximated as a Poisson distribution with $\lambda = np$.

$$Binom(n, p) \rightarrow Poiss(\lambda = np) \tag{3}$$

Can make computations easier (although nowadays, often not as necessary with modern computers).

Means and Variances of Common Discrete Distributions

Means and Variances of Common Discrete Distributions

- $X \sim Bern(p)$

$$E[X] = p \qquad Var[X] = p(1-p)$$

- $X \sim Binom(n, p)$

$$E[X] = np \qquad Var[X] = np(1-p)$$

- $X \sim Poiss(\lambda)$

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

Recall,

$$\begin{aligned} E[X] &= \mu = \sum_i k_i P(X = k_i) \\ Var[X] &= \sigma^2 = \sum_i (k_i - \mu)^2 P(X = k_i) \end{aligned}$$