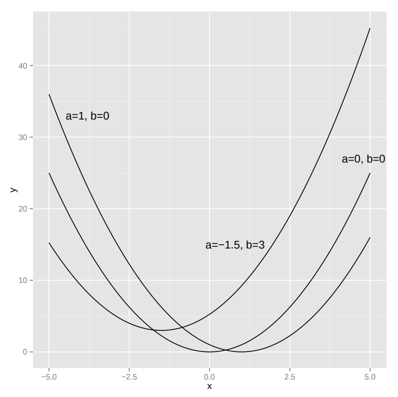
Common Distribution Families

Parametric Distribution Families

The notion of a *parametric family* of distributions is a key concept that will recur throughout the lectures.

Ex: Consider plotting the curves $g_{a,b}(x) = (x - a)^2 + b$. For each a and b, we get a different parabola.



Parametric Function Families

- This is a family of curves, thus a family of functions.
- Numbers a and b are the **parameters** of the family.
- x is not a parameter but an argument of each function.
- The parameters a and b are indexing the curves.

Common Distribution Families - I
Discrete Distributions

Parametric Families of pmfs

Probability mass functions are still functions.

- The domains of these functions are typically the integers.
- They can come in parametric families, indexed by parameters.
- Some parametric families of pmfs have been named due to their usefulness.
- They are famous because they fit real data well in various settings.

- **Note:** Do not assume that we always "must" use pmfs from some family.

Distribution Based on Independent (Bernoulli) Trials

Bernoulli Distribution (with parameter p)

Definition: A random variable X has a *Bernoulli(p)* distribution, denoted $X \sim \text{Bernoulli}(p)$, if its pmf is of the form

$$f_X(x|p) = \begin{cases} p^x(1-p)^{1-x} & \text{for } x = 0 \text{ or } 1, \\ 0 & \text{otherwise,} \end{cases}$$

or

$$f_X(x|p) = p^x(1-p)^{1-x} I_{\{0,1\}}(x)$$

where the parameter p satisfies $0 \leq p \leq 1$.

Representation of Bernoulli Distribution

This is sometimes written as

$$X = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

Notes on Bernoulli Distribution

1. A *Bernoulli trial* (named after James Bernoulli) is an experiment with exactly two possible outcomes.
2. Bernoulli random variable $X = 1$ if "success" occurs and $X = 0$ if "failure" occurs where the probability of success is p (and probability of failure is $1 - p$).

Descriptive Measures for Bernoulli Distribution

Mean: $E[X] = p$,

Variance: $\text{Var}(X) = p(1 - p)$,

MGF: $M_X(t) = pe^t + (1 - p)$.

The Geometric Family of Distributions

Coin Tossing Example: Recall our example of tossing a coin until we get the first head, with N denoting the number of tosses needed.

$$p_N(k) = \left(1 - \frac{1}{2}\right)^{k-1} \cdot \frac{1}{2}, \quad k = 1, 2, ...$$

Here, getting a head is a "success" and a tail is a "failure".

Die Rolling Example: Define M to be the number of rolls of a die needed until the number 5 shows up.

$$p_M(k) = \left(1 - \frac{1}{6}\right)^{k-1} \cdot \frac{1}{6}, \quad k = 1, 2, ...$$

Here, getting a 5 is a "success".

The Independent Bernoulli Trials

Here tosses of the coin and rolls of the die are **Bernoulli trials**.

B_i is 1 for success on the i^{th} trial, 0 for failure, with success probability p. For instance, p is 1/2 for a coin, and 1/6 for a die.

Assumptions for Independent Bernoulli Trials Setting (or Experiment)

1. The experiment consists of a sequence of independent Bernoulli trials.
2. Each trial can result in either a success (S) or failure (F).
3. The probability of success, denoted as p, is constant from trial to trial.

Geometric Distribution (with parameter p)

Consider a sequence of independent Bernoulli trials with probability of success p on each trial (i.e., consider the setting of independent Bernoulli trials).

Definition:
Let X represent the number of trials till the first success. Then X has geometric distribution with parameter p, denoted $X \sim \text{Geometric}(p)$, and has pmf

$$f_X(x|p) = P(X = x|p) = (1-p)^{x-1}\, p \quad \text{for } x = 1, 2, \ldots \text{ and } p \in (0, 1).$$

It is often referred to as a discrete "waiting time" random variable. It represents how long (in terms of the number of trials) one has to wait for one success to occur.

Descriptive Measures

Mean: $E[X] = \frac{1}{p}$,

Variance: $\text{Var}(X) = \frac{1-p}{p^2}$,

cdf: $F(x) = P(X \leq x) = 1 - (1-p)^x$

MGF: $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$.

Example

For $X \sim \text{Geometric}(p)$, show that probability of observing n failures already is $(1-p)^n$ for $n = 1, 2, \ldots$.

Solution "Observing n failures already (i.e., n failures already occurred)" is equivalent to "observing first success at $n+1$)st trial or at a later trial": That is, $X > n$ (i.e., $X \geq n$). So,

$$\begin{aligned}
P(X > n) &= P(X = n+1) + P(X = n+2) + \ldots \\
&= (1-p)^n \times p + (1-p)^{n+1} \times p + \ldots \\
&= (1-p)^n \times p[1 + (1-p) + (1-p)^2 \ldots] \\
&= (1-p)^n \times p \sum_{k=0}^{\infty} (1-p)^k
\end{aligned}$$

Using the formula for the sum of an infinite geometric series:
$S = \sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$, we get:
$P(X > n) = (1-p)^n \times p \times \frac{1}{1-(1-p)} = (1-p)^n \times p \times \frac{1}{p} = (1-p)^n$.

Memoryless Property of Geometric Distribution

This distribution has an interesting property, known as the "memoryless property". For integers $s, t \geq 1$,

$$P(X > t + s | X > t) = P(X > s).$$

Probability of getting s more failures, having already observed t failures, is the same as the probability of observing s failures from the start.

That is, this distribution "forgets" what has occurred before.

Why?

For $s, t \geq 1$

$$\begin{aligned} P(X > t + s | X > t) &= \frac{P(X > t + s \text{ and } X > t)}{P(X > t)} \\ &= \frac{P(X > t + s)}{P(X > t)} = \frac{(1-p)^{t+s}}{(1-p)^t} \\ &= (1-p)^s = P(X > s). \end{aligned}$$

Note: In fact, geometric r.v. is the only positive discrete r.v. that is memoryless.

Example: Failure Times

The geometric distribution is sometimes used to model "lifetimes" or "time until failure" of components (in industry). If the probability is .001 that a light bulb will fail on any given day, then find the probability that it will last at least 30 days from now.

Solution:

Let $X = \#$ of days till the light bulb fails for the first time. So, here "success" is the failure of light bulb. Then $X \sim \text{Geometric}(p = .001)$ and the probability is:

$$P(X > 30) = (.999)^{30} = 0.970.$$

Example: Aging of People

Suppose X is the number of years one lives. Then

$$P(\text{s/he lives two more years}) = P(X > \text{current age} + 2 \mid X > \text{current age}) = P(X > 2).$$

Clearly, this model is not realistic for humans. The remaining life time changes as people get older.

Caveat: Geometric distribution is not applicable to modeling lifetimes for which the probability of failure is expected to increase over time. There are other distributions for modeling aging problems.

R Functions for Geometric Distribution

You can simulate geometrically distributed random variables in R.

R Functions: The R functions for a geometrically distributed random variable X with success probability p:

- `dgeom(i,p)`: $P(X = i)$
- `pgeom(i,p)`: $P(X \leq i)$
- `qgeom(q,p)`: find c such that $P(X \leq c) = q$
- `rgeom(n,p)`: generate n variates

Important Note on R's Definition:
R and some other software define geometric distributions as the number of failures before the first success.

```
> dgeom(2,0.4)
[1] 0.144
```

The Binomial Family of Distributions

A geometric distribution arises when we have independent Bernoulli trials with parameter p, with a variable number of trials (N) but a fixed number of successes (which is 1).

The **binomial distribution** arises when we have the opposite — a fixed number of independent Bernoulli trials (n) but a variable number of successes (say X).

Example: For example, say we toss a coin five times, and let X be the number of heads we get. We say that X is binomially distributed with parameters $n = 5$ and $p = 1/2$. The probability $P(X = 2)$ can be calculated as:

$$P(X = 2) = \binom{5}{2} 0.5^2 (1 - 0.5)^3 = \binom{5}{2} / 32 = 5/16$$

Binomial Distribution

Definition: A random variable X has a *Binomial*(n, p) distribution, denoted $X \sim \text{Binomial}(n, p)$, if its pmf is of the form

$$f_X(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x},$$

$$\text{for } x = 0, 1, \ldots, n, \text{ with } n > 0, \text{ and } 0 \le p \le 1,$$

or

$$P(X = x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\ldots,n\}}(x).$$

So, again we have a parametric family of distributions, in this case a family having two parameters, n and p.

Representation using Bernoulli Variables

Let's write X as a sum of those 0-1 Bernoulli variables:

$$X = \sum_{i=1}^{n} B_i$$

where B_i is 1 or 0, depending on whether there is success on the i^{th} trial or not, more precisely, $B_1, \ldots, B_n$ are independent identical Bernoulli r.v.'s.

Descriptive Measures for Binomial Distribution:

Using properties of $E()$ and $Var()$:

Mean: $E[X] = np$,

Variance: $\text{Var}(X) = np(1-p)$,

MGF: $M_X(t) = (pe^t + (1-p))^n$.

Example: Sampling with Replacement

Consider sampling **with replacement** from an urn containing n items, m of which are defective. Let X represent the number of defective items in a sample of size k. The probability that a sample of size k contains x defectives is:

$$P(X = x) = f_X\left(x|k, p = \frac{m}{n}\right) = \binom{k}{x}\left(\frac{m}{n}\right)^x\left(1 - \frac{m}{n}\right)^{k-x},$$

$$\text{for } x = 0, 1, \ldots, k.$$

Example: Defective Screws (Ross, 1988)

The screws produced by a certain company are defective with probability 0.01 independently. The company sells screws in packages of 10 and offers a money-back guarantee for packages with 2 or more defective screws. What proportion of packages sold will the company replace?

Solution: The question is equivalent to "What is the probability that a package has to be replaced." Notice that a package has to be replaced if 2 or more are defective in it. Let X be the number of defective screws in a package, then we want to find $P(X \geq 2)$. Also, $X \sim \text{Binomial}(10, .01)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{10}{0} 0.01^0\, 0.99^{10} - \binom{10}{1} 0.01^1\, 0.99^9 \\ &\approx 1 - 0.9044 - 0.0914 \approx .004. \end{aligned}$$

R Functions for Binomial Distribution

Relevant functions for a binomially distributed random variable X for k trials and with success probability p are:

- **dbinom(i,k,p)**: $P(X = i)$
- **pbinom(i,k,p)**: $P(X \leq i)$
- **qbinom(q,k,p)**: smallest c such that $P(X \leq c) \geq q$
- **rbinom(n,k,p)**: generate n independent values of X

The Negative Binomial Family of Distributions

Recall the geometric distribution arises as N, the number of tosses of a coin needed to get the first head. Now generalize to N being the number of tosses needed to get the r^{th} head, where r is fixed.

Understanding the "Prob. of $N = k$" for the case $r = 3$, $k = 5$: $\{N = 5\} = \{2 \text{ heads in the first } 4 \text{ tosses and head on the } 5^{th} \text{ toss}\}$ The event $2 \text{ heads in the first } 4 \text{ tosses}$ is a binomial probability:

$$P(2 \text{ heads in the first } 4 \text{ tosses}) = \binom{4}{2}\left(\frac{1}{2}\right)^4$$

So, by independence,

$$P(N = 5) = P(2 \text{ heads in the first } 4 \text{ tosses}) \times P(\text{head on the } 5^{th} \text{ toss})$$

$$P(N = 5) = \binom{4}{2}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right) = \binom{4}{2}\left(\frac{1}{2}\right)^5 = \frac{3}{16}$$

Probability of $N = k$

The negative binomial distribution family, indexed by parameters r and p, counts the number of independent trials with success probability p needed until we get r successes.

Probability Mass Function

$$p_N(k) = P(N = k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r, \quad k = r, r+1, ... \quad (1)$$

This can be expressed as:

$$N = G_1 + ... + G_r = \sum_{i=1}^{r} G_i$$

where each G_i is the number of tosses between the successes numbers $i - 1$ and i. Note that G_i's are independent and $G_i \sim Geo(p)$.

Descriptive Measures:

For $X \sim NB(p, r)$, we have

Mean: $E[X] = \dfrac{r}{p}$

Variance: $\mathrm{Var}(X) = \dfrac{r(1-p)}{p^2}$

cdf: $F(x)$ has no closed form.

MGF: $M_X(t) = \left(\dfrac{pe^t}{1-(1-p)e^t}\right)^r = \left(\dfrac{p}{e^{-t}-(1-p)}\right)^r.$

R Functions for Negative Binomial Distribution

Relevant functions for a negative binomially distributed random variable X with success parameter p are:

- **dnbinom(i,size=r,prob=p)**: $P(X = i)$
- **pnbinom(i,size=r,prob=p)**: $P(X \leq i)$
- **qnbinom(q,size=r,prob=p)**: smallest c such that $P(X \leq c) = q$
- **rnbinom(n,size=r,prob=p)**: generate n independent values of X

Example: Backup Batteries

A machine contains one active battery and two spares. Each battery has a 0.1 chance of failure each month. Let L denote the lifetime of the machine, i.e. the time in months until the third battery failure. Find $P(L = 12)$.

Solution:

The number of months until the third failure has a negative binomial distribution, with $r = 3$ and $p = 0.1$. Thus the answer is obtained by Equation (1), with $k = 12$:

$$P(L = 12) = \binom{11}{2}(1 - 0.1)^9 0.1^3 = 0.0213$$

Sample Histograms

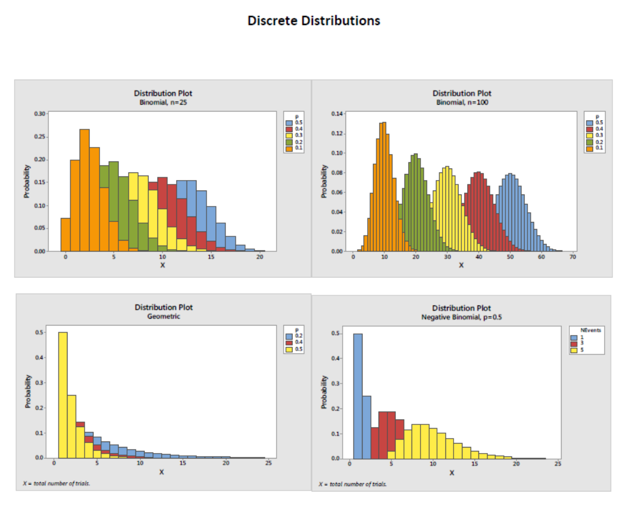


Figure 1: Sample histograms of the pmfs for various discrete distributions.

The Poisson Family of Distributions

The Poisson Distribution

The **Poisson Distributions** differ from the geometric, binomial, and negative binomial families, which have clear qualitative descriptions of their origins. Instead, the Poisson distribution is primarily used to model random events where count of some quantity over a period of time, space, region, length, volume, etc. is the variable of interest.

Some Examples:

- the number of typos on a page of a book,
- the number of incoming 911 calls at a police switchboard in one day,
- the number of α-particles discharged in a fixed period of time from some radio active material,
- the # of customers entering a credit union on a given day,
- the number of defects per square yard of a certain fabric, etc.

Notes on Poisson Distribution

1. A Poisson distribution is typically used to model the probability distribution of the number of occurrences per unit time or per unit area (with λ being the *intensity rate*).
2. *The basic assumption:* the probability of an event occurring is proportional to the length of the time interval.
3. *A useful result (for verifying the descriptive measures below):* By Taylor series expansion, we have $e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}$

The pmf for the Poisson distribution is:

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

Descriptive Measures:

- *Mean:* $E[X] = \lambda$
- *Variance:* $\text{Var}(X) = \lambda$
- *MGF:* $M_X(t) = e^{\lambda(e^t - 1)}$

Example: # of Calls

If there are 6 calls in 3 minutes on average at a call center, what is the probability that:

(a) there will be no calls in the next minute?
(b) at least two calls in the next minute?

Solution: Let X = number of calls in a minute, then X has a Poisson distribution with $E[X] = \lambda = 6/3 = 2$, thus

$$\text{(a) } P(X = 0) = \frac{e^{-2}(2)^0}{0!} \approx 0.135,$$

$$\begin{aligned}\text{(b) } P(X \geq 2) &= 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - 0.135 - \frac{e^{-2}(2)^1}{1!} \\ &= 1 - 0.189 - 0.271 \approx 0.54.\end{aligned}$$

R Functions for Poisson Distribution

Relevant R functions for a Poisson distributed variable X with parameter λ:

- **dpois(i,lambda)**: $P(X = i)$
- **ppois(i,lambda)**: $P(X \leq i)$
- **qpois(q,lambda)**: smallest c such that $P(X \leq c) = q$
- **rpois(n,lambda)**: generate n independent values of X

Example: Broken Rod

Recalling the broken glass rod example, suppose now that the number of breaks is random. A potential model is Poisson. We model the number of pieces minus 1 (break points) as Poisson.

Suppose we wish to find the expected value of the shortest piece, via simulation. The code is similar to that in Section 2.6, but we must first generate the number of break points (see below code).

```
minpiecepois <- function(lambda) {
   nbreaks <- rpois(1,lambda)
   breakpts <- sort(runif(nbreaks))
   lengths <- diff(c(0,breakpts,1))
   min(lengths)
}
bkrodpois <- function(nreps,lambda,q) {
   minpieces <- replicate(nreps,minpiecepois(lambda))
   mean(minpieces < q) }
> bkrodpois(10000,5,0.02)
[1] 0.4655
```

The Power Law Family of Distributions

The Power Law Distribution

The Power Law family has gained attention recently due to its application in random graph models.

The Model: The probability mass function is given by:

$$p_X(k) = ck^{-\gamma},\ \ k = 1, 2, 3, ...$$

with the condition $\gamma > 1$ to ensure the sum of probabilities isn't infinite.

The value of c is set to ensure the sum is 1.0:

$$1.0 = \sum_{k=1}^{\infty} ck^{-\gamma} \approx c \int_{1}^{\infty} k^{-\gamma}\ dk = c/(\gamma - 1)$$

Thus, $c \approx \gamma - 1$.

Applications and Interests

The Power Law family is traditional, but has seen renewed interest. Real-world social networks often exhibit power law behavior in degree distributions.

A study of the Web found:

- Number of incoming links to a page: $\gamma = 2.1$
- Number of outgoing links from a page: $\gamma = 2.7$

Interest also arises from the **fat tails** of power laws. Extreme values, or **black swans**, are more probable under a power law than a normal distribution with the same mean.

Power Law with Exponential Cutoff

A variant of the power law is combined with a geometric distribution:

$$p_X(k) = ck^{-\gamma}q^k$$

This is a two-parameter model, with parameters γ and q. c is chosen to ensure the pmf sums to 1.0.

This model fits some data better than the pure power law, but its tail decays exponentially in k.

Real Data and Power Laws: Some real data sets fit well with power laws, others do not.

Reference: Power-Law Distributions in Empirical Data by A. Clauset, C. Shalizi, and M. Newman.

The paper also discusses methods to estimate γ.

Recognizing Some Parametric Distributions When You See Them

Three of the discrete distribution families considered arise in specific settings involving independent trials.

Distributions from Independent (Bernoulli) Trials

- The binomial family: Distribution of the number of successes in a fixed number of trials.
- The geometric family: Distribution of the number of trials needed to obtain the first success.
- The negative binomial family: Distribution of the number of trials needed to obtain the kth success.

These scenarios are common, making these distribution families particularly noteworthy.

Distributions Without Underlying Structure

The Poisson and power law distributions:

- Lack a specific underlying structure like the previous distributions.
- Are renowned because they often fit well to many real data sets.

For the binomial, geometric, and negative binomial distributions, the fundamental nature of the setting implies the distribution.

Recognition is Key! *You should make a strong effort to recognize these settings automatically when you encounter them.*

Example: Analysis of Social Networks

One of the simplest models of social networks was developed by Erdös and Renyi. Consider:

- n people (or Web sites, etc.).
- $\binom{n}{2}$ potential links (undirected graph).
- Each pair of people has a link with probability p.
- Each pair of people does not have a link with probability $1 - p$.
- All pairs are independently having or not having links.

Degree Distribution

The degree distribution D_i for a node i is binomial with parameters n-1 and p.

Distribution of Links for k Nodes: Consider k nodes, 1 through k, among n total nodes. Let T be the number of links involving these nodes. The distribution of T is binomial. In particular, with $k = 4$ and $n = 9$:

- $\binom{4}{2} = 6$ potential links among the special nodes.
- 4 special nodes, each with $9 - 4 = 5$ potential links to the outside.
- Total of 26 potential links.

Binomial Distribution of T

The distribution of T is binomial with:

$$k(n-k)+\binom{k}{2} \tag{2}$$

trials and success probability p.

That is, $T \sim \text{Binomial}(n_t, p)$ where $n_t = k(n-k)+\binom{k}{2}$.