Statistical Inference Notes

Liam Flaherty

May 31, 2023

Contents

1 Set Theory And Basic Probability

1.1 Definitions

Definition 1.1. Sample Space (Ω): the set of all possible outcomes in an experiment. For example, the sample space for the outcome of rolling two fair dice is $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$.

Definition 1.2. Probability Measure (P): A function $P : (A \subset \Omega) \to [0, 1]$ satisfying $P(\Omega) = 1$, $A \subset \Omega \implies P(A) \geq 0$, and A_1, A_2 disjoint $\implies P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

Definition 1.3. Combination: The number of ways one can select r elements from a set of n when the order of the selection doesn't matter. When the elements are **not replaced** after each selection, the number of combinations is the binomial coefficient $\binom{n}{r} = \frac{n!}{(n-r)!r!}$. When the elements **are replaced** after each selection, the number of combinations is $\binom{n+r-1}{r}$. For example, the number of ways a company can divide 10 people into three groups of size three, three, and four is $\binom{10}{4}\binom{6}{3} = \frac{10!}{6!4!}\frac{6!}{3!3!} = \frac{10!}{4!3!3!} = 4200$ since the order in which the people are selected to be in a group doesn't matter (combination), since a person can't be in two groups (combination without replacement), and since after selecting for the first group there is a remaining set of 6 to select the second group from, after which the third group is determined.

Definition 1.4. Permutation: The number of ways one can select r elements from a set of n when the order of the selection matters. When the elements are **not replaced** after each selection, the number of permutations is $r!\binom{n}{r} = \frac{n!}{(n-r)!}$. When the elements **are replaced** after each selection, the number of permutations is n^r. For example, the number of four letter "words" is 26^4 while the number of assortments of gold, silver, and bronze medal winners in a competition of 10 is $\frac{10!}{7!} = 10 \cdot 9 \cdot 8 = 720$.

Definition 1.5. Marginal Probability, $P(A)$: The probability an event in the sample space occurs, without conditioning on another event. For example, the probability of rolling a five with a fair dice is $\frac{1}{6}$.

Definition 1.6. Conditional Probability, $P(A|B)$: The probability an event A in the sample space occurs, given that the event B has occurred. In general, $P(A|B) = \frac{P(A \cap B)}{P(B)}$. **Bayes' Theorem**, $P(A|B) = \frac{P(A)}{P(B)}P(B|A)$, may be useful in situations where we have some conditional probabilities but not others. For example, the probability of rolling a total greater than 9 with the role of two fair dice, *with the knowledge that your first roll is a five*, is $\frac{1}{3}$.

Definition 1.7. Independence: Events A and B are independent provided the conditional probability of A given B is the marginal probability of A; $P(A \cap B) = P(A)P(B)$. For example, the probability one is six feet tall is likely independent of the probability that one catches COVID.

Definition 1.8. Disjoint (Mutually Exclusive): Events A and B are disjoint provided A and B cannot both occur; $P(A \cap B) = 0$. For example, one cannot be both six feet tall and five feet tall.

1.2 Theorems And Further Examples

> mytable

	Combin. W/O Replacement	Combin. W/ Replacement	Perm. W/O Replacement	Perm. W/ Replacement
1	ab	aa	ab	aa
2	ac	ab	ac	ab
3	ad	ac	ad	ac
4	ae	ad	ae	ad
5	bc	ae	ba	ae
6	bd	bb	bc	ba
7	be	bc	bd	bb
8	cd	bd	be	bc
9	ce	be	ca	bd
10	de	cc	cb	be
11	''	cd	cd	ca
12	''	ce	ce	cb
13	''	dd	da	cc
14	''	de	db	cd
15	''	ee	dc	ce
16	''	''	de	da
17	''	''	ea	db
18	''	''	eb	dc
19	''	''	ec	dd
20	''	''	ed	de
21	''	''	''	ea
22	''	''	''	eb
23	''	''	''	ec
24	''	''	''	ed
25	''	''	''	ee
26	''	''	''	''
27	Binom Coeff (n k): P/k!=n!/(n-k)!k!	Binom Coeff: (n+k-1 k)	Gen Formula: Perm=n!/(n-k)!	Gen Formula: n^k
28	5!/[(5-2)!2!]=10	(5+2-1)!/[(5-2)!2!]=15	5!/[(5-2)!2!]=20	5^2=25

Figure 1.1: Select Two Objects From A Set Of Five

```
#####combinatorics#####
selectset=5                     #example, number of items to choose from#
selectchoose=2                  #example, number of items to select#


###In combinations, order doesn't matter, i.e. (a,b)=(b,a)###
a=data.frame(                   #without replacement, how many different ways can you select r objects from a set of size n?#
  combinations(
    n=selectset, r=selectchoose,
    v=letters[1:selectset], repeats.allowed=F))
a=data.frame(paste0(a[,1],a[,2]))
colnames(a)="Combin. W/O Replacement"
dim(a)                           #gives size of dataframe in terms of x rows by y columns#
choose(selectset,selectchoose)   #can also use binomial coefficient formula (n choose k function) to get size#
a                                #general formula is P/k! or [n!/((n-k)!k!)]#


b=data.frame(                   #with replacement, how many different ways can you select r objects from a set of size n?#
  combinations(
    n=selectset, r=selectchoose,
    v=letters[1:selectset], repeats.allowed=T))
b=data.frame(paste0(b[,1],b[,2]))
colnames(b)="Combin. W/ Replacement"
dim(b)
choose(selectset+selectchoose-1, selectchoose) #can again use binomial coefficient formula (n choose k function) to get size#
b                                              #general formula is similar to above, [(n+k-1)!/((n-1)!k!)]#
```

Figure 1.2: R Script To Get Combinations

1.3 Problems

1.73) A system has n independent units, each of which fails with probability p. The system fails only if k or more of the units fail. What is the probability that the system fails?

The probability the first k units fail is p^k, so we can already give an unambitious lower bound. The probability that *exactly* k units fail is calculated by multiplying the probability that k units fail (p^k) by the probability that $n-k$ units don't fail ($(1-p)^{n-k}$) by the number of ways that exactly k units can fail (call it m).

Observe we can count the number of ways that exactly k units can fail by using the binomial coefficient since doing so is equivalent to selecting k objects from a set of size n without replacement. So we can improve the bound to $\binom{n}{k}(p)^k(1-p)^{n-k}$, which is the probability that exactly k units fail.

We also need to account for the probability that exactly $k+1$ units fail, exactly $k+2$ units fail, ..., and all n of the units fail. With the same reasoning as above, we see:

$$\begin{aligned} P(k \text{ units fail}) &= \binom{n}{k}(p)^k(1-p)^{n-k} \\ P(k+1 \text{ units fail}) &= \binom{n}{k+1}(p)^{k+1}(1-p)^{n-k-1} \\ P(k+2 \text{ units fail}) &= \binom{n}{k+2}(p)^{k+2}(1-p)^{n-k-2} \\ &\vdots \quad \vdots \\ P(n \text{ units fail}) &= \binom{n}{n}(p)^n(1-p)^0 = p^n \end{aligned}$$

So we see that the probability the system fails is $\sum\limits_{j=k}^{n} \binom{n}{j}(p)^j(1-p)^{n-j}$

1.33) An elevator containing five people can stop at any of seven floors. What is the probability that no two people get off at the same floor? Assume that the occupants act independently and that all floors are equally likely for each occupant.

The first person has seven floors to choose from, the second person also has seven floors to choose from, etc. So the total number of ways that the five people can get off is 7^5. There are 7 choices for the first person, and after the first person makes their choice, the second person has six choices to ensure they don't get off on the same floor, etc. So the probability that no two people get off on the same floor is $\frac{\prod_{i=3}^{7} i}{7^5} = \frac{\prod_{i=3}^{6} i}{7^4} = \frac{360}{2401}$.

2 Random Variables

2.1 Definitions

Definition 2.1. Random Variable: A function $X : \Omega \to \mathbb{R}$. Random variables can be categorized as either **discrete** (when the random variable can take on a finite or at most countably infinite number of values) or **continuous**. For example, if a coin is flipped twice the sample space is $\{HH, HT, TH, TT\}$, and a possible random variable could be the total number of heads in the observation.

Definition 2.2. Support: The set of values which a random variable can take on (its image). For example, if a random variable is the total number of heads minus the total number of tails after two coin flips, the support of the random variable is $\{2, 0, -2\}$.

Definition 2.3. Probability Mass Function (PMF): A function $p : \mathbb{R} \to [0, 1]$ defined by $p_X(x_i)=P(X=x_i)$ where X is a **discrete random variable** and x_i is a member of the support. For example, if an experiment is flipping a fair coin twice and the random variable gives the difference between the number of heads and tales, the PMF is $p(x) = \begin{cases} 1/4, & x = \pm 2 \\ 1/2, & x = 0 \end{cases}$

Definition 2.4. Probability Density Function (PDF): A piece-wise continuous function $f : \mathbb{R} \to [0, 1]$ such that $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\, dx = 1$ (where x is in the support of a **continuous random variables** X). Probabilities for continuous random variables must be given on an interval; the PDF is given by $P(a \leq X \leq b) = \int_a^b f(x)\, dx$. For example, if an experiment measures the number of years a person lives, the random variable could give the probability a person dies between a and b years.

Definition 2.5. Cumulative Distribution Function (CDF): A monotone increasing function $F : \mathbb{R} \to [0, 1]$ defined by $F(x)=P(X \leq x)$ where $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$. In contrast to PMF's or PDF's, which are often denoted by lowercase letters, CDF's are usually denoted with capital letters. They can be used to find the **$\mathbf{p^{th}}$ Quantile**, which is the value x_p such that $F(x_p) = p$.

Definition 2.6. Expected Value ($\mathbb{E}(X)$ or μ_X): Informally, the mean of a large number of independent outcomes of a random variable. Expected value is the first moment. If X is a discrete random variable with probability mass function p and support x_i, then the expected value is given by $\sum\limits_{i \in \mathbb{N}} x_i \cdot p(x_i)$ (provided the sum converges). If X is a continuous random variable with probability density function f, then the expected value is given by $\int_{-\infty}^{\infty} x \cdot f(x)\, dx$ (provided the integral converges).

Definition 2.7. Variance ($\mathbb{V}(X)$ or σ_X^2): Informally, the dispersion of a random variable. Variance is the second central moment. Where $\mathbb{E}(X)$ is the expected value of a random variable X, the variance is $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. The **standard deviation** is the square root of variance, and is useful because it is in the same units as the random variable.

Definition 2.8. Moment Generating Function (MGF): $M(t) = \mathbb{E}(e^{tX})$. The k^{th} derivative of the MGF evaluated at zero gives the k^{th} **raw moment**, while the k^{th} derivative of the MGF evaluated at the mean gives the k^{th} **central moment**.

2.2 Theorems And Further Examples

Theorem 2.1. Expected Value Is Linear: For scalars a and b, $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$

Proof.

$$\begin{aligned}
\mathbb{E}(aX+b) &= \int_{-\infty}^{\infty} (ax+b) f_X(x)\,dx = \int_{-\infty}^{\infty} ax f_X(x) + b f_X(x)\,dx \\
&= \int_{-\infty}^{\infty} ax f_X(x)\,dx + \int_{-\infty}^{\infty} b f_X(x)\,dx \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} f_X(x)\,dx \\
&= a\mathbb{E}(X) + b
\end{aligned}$$

■

Theorem 2.2. Variance Is Translation Invariant: For scalars a and b, $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$

Proof.

$$\begin{aligned}
\mathbb{V}(aX+b) &= \mathbb{E}\Big[\big((aX+b) - \mathbb{E}(aX+b)\big)^2\Big] \\
&= \mathbb{E}\Big[\big(aX+b-a\mathbb{E}(X)-b\big)^2\Big] = \mathbb{E}\Big[\big(aX - a\mathbb{E}(X)\big)^2\Big] \\
&= \mathbb{E}\Big[\big(a^2X^2 - 2a^2\mathbb{E}(X)X + a^2\mathbb{E}(X)^2\big)\Big] = \mathbb{E}\Big[a^2\big(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2\big)\Big] \\
&= a^2\mathbb{E}\Big[\big(X - \mathbb{E}(X)\big)^2\Big] = a^2\mathbb{V}(X)
\end{aligned}$$

■

Theorem 2.3. Markov's Inequality: For any positive random variable X where expectations exist, $P(X \geq t) \leq \mathbb{E}(X)/t$.

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{t} x f_X(x)\,dx + \int_{t}^{\infty} x f_X(x)\,dx \\
&\geq \int_{t}^{\infty} x f_X(x)\,dx \geq \int_{t}^{\infty} t f_X(x)\,dx \\
&\geq t\int_{t}^{\infty} f_X(x)\,dx = t \cdot P(X \geq t)
\end{aligned}$$

■

Theorem 2.4. Chebyshev's Inequality: For any $\varepsilon > 0$, $P(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \mathbb{V}(X)/\varepsilon^2$.

Proof.

$$\begin{aligned}
&P\big(|X-\mathbb{E}(X)| \geq \varepsilon\big) = P\big((X-\mathbb{E}(X))^2 \geq \varepsilon^2\big) \quad \text{Apply Markov Inequality to } \big(X-\mathbb{E}(X)\big)^2 \\
&\leq \mathbb{E}\Big(\big(X-\mathbb{E}(X)\big)^2\Big)/\varepsilon^2 = \mathbb{V}(X)/\varepsilon^2
\end{aligned}$$

■

2.3 Problems

2.31) Phone calls are received at a certain residence as a Poisson process with parameter $\boldsymbol{\lambda = 2}$ per hour.

a. If Diane takes a 10-min shower, what is the probability that the phone rings during that time?

The pmf for a Poisson process is $P(X{=}k){=}\frac{\lambda^k e^{-\lambda}}{k!}$. The probability the phone rings in any 10 minute interval is the complement of the probability the phone doesn't ring in a 10 minute interval. The probability that zero calls come in is given by the parameter $\lambda = 10 \cdot \frac{2}{60}$, so the probability the phone rings while Diane is in the shower is $1 - P(X = 0) = 1 - e^{-\frac{1}{3}} \approx 28.35\%$.

b. How long can her shower be if she wishes the probability of receiving no phone calls to be at most .5?

The probability the phone doesn't ring in a t minute interval is given by $P(X = 0) = e^{-t\frac{1}{30}}$. We want to maximize t where $e^{-t\frac{1}{30}} < 0.5$. Then we need $t < -30\ln(0.5) \approx 20.79$ minutes.

2.40) Suppose that $\boldsymbol{X}$ has the density function $\boldsymbol{f(x) = cx^2}$ for $\boldsymbol{0 \leq x \leq 1}$ and $\boldsymbol{f(x) = 0}$ otherwise. Find $\boldsymbol{c}$, find the cumulative distribution function, and determine $\boldsymbol{P(.1 \leq X \leq .5)}$

The probability density function is the function f such that $F_X(x)$, the cumulative distribution function, is equal to $\int_{-\infty}^{x} f(t)\,dt$. Here $f(x) = 0$ for $x < 0$, so $F_X(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x f(t)\,dt = \int_0^x ct^2\,dt = \frac{c}{3}t^3 \mid_0^x = \frac{c}{3}x^3$ for values of $x \in [0,1]$, while $F_X(x) = 0$ for $x < 0$ and $F_X(x) = 1$ for $x > 1$.

Since f only gives non-zero probabilities in the interval $[0,1]$, the integral of the pdf over this interval must be 1, i.e. $F_X(x){=}\int_0^1 f(x)\,dx{=}\frac{c}{3}x^3 \mid_0^1 = \frac{c}{3} = 1$ and so $c = 3$.

As X is a continuous random variable, to find $P(.1 \leq X \leq .5)$, we can integrate the pdf with the limits 0.1 and 0.5. See $\int_{0.1}^{0.5} f(x)\,dx{=}\frac{c}{3}x^3 \mid_{0.1}^{0.5} = \frac{\boxed{3}}{3}x^3 \mid_{0.1}^{0.5} {=} 0.5^3 - 0.1^3 = 0.124$

4.5) Let X have the density $f(x) = \frac{1+\alpha x}{2}$ for $x, \alpha \in [-1,1]$. Find $\mathbb{E}(X)$ and $\mathbb{V}(X)$.

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\,dx = \int_{-1}^{1} x \cdot f(x)\,dx \\ &= \frac{1}{2}\int_{-1}^{1} x + \alpha x^2\,dx = \frac{1}{2}\left[\left(\frac{x^2}{2} + \frac{\alpha x^3}{3}\right)|_{-1}^{1}\right] = \frac{1}{2}\left[\left(\frac{1}{2}+\frac{\alpha}{3}\right) - \left(\frac{1}{2}-\frac{\alpha}{3}\right)\right] = \frac{\alpha}{3}\end{aligned}$$

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \left(\frac{\alpha}{3}\right)^2 = \left[\frac{1}{2}\int_{-1}^{1} x^2 + \alpha x^3\,dx\right] - \left(\frac{\alpha}{3}\right)^2 \\ &= \left[\frac{1}{2}\int_{-1}^{1} x^2 + \alpha x^3\,dx\right] - \left(\frac{\alpha^2}{9}\right) = \frac{1}{2}\left[\left(\frac{x^3}{3} + \frac{\alpha x^4}{4}\right)|_{-1}^{1}\right] - \left(\frac{\alpha^2}{9}\right) \\ &= \frac{1}{2}\left[\left(\frac{1}{3}+\frac{\alpha}{4}\right) - \left(\frac{-1}{3}+\frac{\alpha}{4}\right)\right] - \left(\frac{\alpha^2}{9}\right) = \frac{1}{2}\left(\frac{2}{3}\right) - \frac{\alpha^2}{9} = \frac{3-\alpha^2}{9}\end{aligned}$$

1) Let X have the density $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{x})^{\alpha-1}e^{-\beta/x}$ whenever $x > 0$ for positive values of α and β. Determine $\mathbb{E}(X^r)$ to calculate the mean and variance of X.

Note that since f is a valid density, $\int_{-\infty}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)}(\frac{1}{x})^{\alpha+1}e^{-\beta/x}\,dx{=}1 \implies \int_{-\infty}^{\infty}(\frac{1}{x})^{\alpha+1}e^{-\beta/x}\,dx{=}\frac{\Gamma(\alpha)}{\beta^\alpha}$ as α and β are constants that don't depend on the integrating variable. Then observe:

$$\begin{aligned}\mathbb{E}(X^r) &= \int_{-\infty}^{\infty} x^r \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{x}\right)^{\alpha+1} e^{-\beta/x}\,dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_{-\infty}^{\infty} x^r \left(\frac{1}{x}\right)^{\alpha+1} e^{-\beta/x}\,dx && \text{Properties of exponents} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)}\int_{-\infty}^{\infty} x^{-(\alpha-r+1)} e^{-\beta/x}\,dx && \text{Same kernel as the note above} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha-r)}{\beta^{\alpha-r}} = \beta^r \frac{\Gamma(\alpha-r)}{\Gamma(\alpha)}\end{aligned}$$

The mean is the first moment, so (recalling that $\Gamma(a+1) = a\Gamma(a)$):

$$\mathbb{E}(X) = \mathbb{E}(X^1) = \beta^{[1]}\frac{\Gamma(\alpha-[1])}{\Gamma(\alpha)} = \beta\frac{\Gamma(\alpha-1)}{\Gamma(\alpha-1+1)} = \beta\frac{\Gamma(\alpha-1)}{(\alpha-1)\Gamma(\alpha-1)} = \frac{\beta}{\alpha-1}$$

The variance is the second central moment, so:

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \beta^{[2]}\frac{\Gamma(\alpha-[2])}{\Gamma(\alpha)} - \frac{\beta^2}{(\alpha-1)^2} = \frac{\beta^2\Gamma(\alpha-2)}{\Gamma(\alpha-1+1)} - \frac{\beta^2}{(\alpha-1)^2} \\ &= \frac{\beta^2\Gamma(\alpha-2)}{(\alpha-1)\Gamma(\alpha-1)} - \frac{\beta^2}{(\alpha-1)^2} = \frac{\beta^2}{(\alpha-1)(\alpha-2)} - \frac{\beta^2}{(\alpha-1)^2} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}\end{aligned}$$

3 Common Distributions

3.1 Discrete Random Variables

1) Bernoulli Distribution: The random variable takes on values either 1 (success) or 0 (failure) with probability p and $1-p$ respectively. For example, a coin is flipped that lands on heads with probability p. The random variable takes value 1 if the coin lands on heads, and 0 if the coin lands on tails.

Notation and Parameter(s): $X \sim \text{Bern}(p)$ Support: $k \in \{0,1\}$

PMF: $p(k) = \begin{cases} p^k(1-p)^k, & x \in \{0,1\} \\ 0, & x \notin \{0,1\} \end{cases}$ CDF: $F(k) = \begin{cases} 0, & k < 0 \\ 1-p, & 0 \leq k < 1 \\ 1, & k \geq 1 \end{cases}$

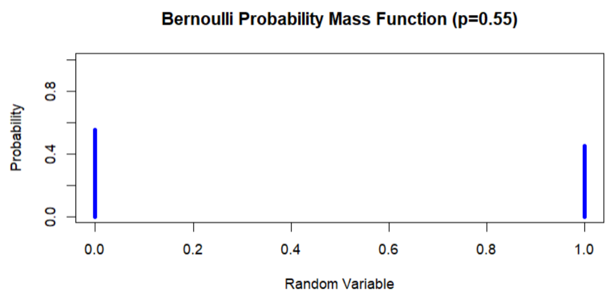

Figure 3.1: Bernoulli PMF Example

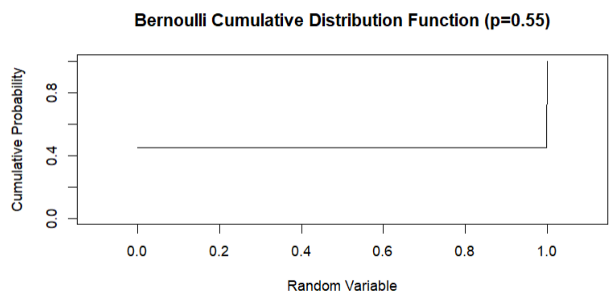

Figure 3.2: Bernoulli CDF Example

```
######Bernoulli Distribution#####
###Parameters###
p=0.55                              #the probability of success#

###PMF###
x=seq(from=0, to=1, by=1)
y=dbern(x,p)                        #use Rlab package#
plot(x, y, type="h",                #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     lwd=5, col="blue",
     ylab="Probability", xlab="Random Variable (Success Or Failure)",
     main=paste0("Bernoulli Probability Mass Function (p=", p, ")"))

###CDF###
x=seq(from=0, to=1, by=0.001)
y=ifelse(x<1, 1-p, 1)               #pbern doesn't work as intended#
plot(x, y, type="s",                #s for step, l for line#
     xlim=c(min(x)-0.1,max(x)+0.1), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable (Success Or Failure)",
     main=paste0("Bernoulli Cumulative Distribution Function (p=", p, ")"))

###Simulations And Questions###
mysim=rbern(100, p)                 #100 random simulations from Bernoulli#
```

Figure 3.3: Bernoulli R Script

2) Binomial Distribution: The random variable is the total number of successes after n independent Bernoulli Trials. For example, the random variable could be the number of hits a baseball player with a "true" batting average of 0.3 gets in 5 at bats.

Notation and Parameter(s): $X \sim \text{Binom}(n, p)$ Support: $k \in \{0, 1, \ldots, n\}$

PMF: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ CDF: $F(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$

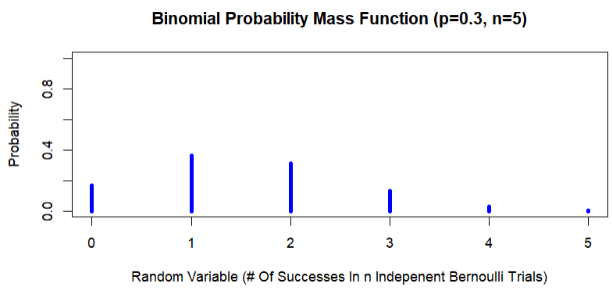

Figure 3.4: Binomial PMF Example

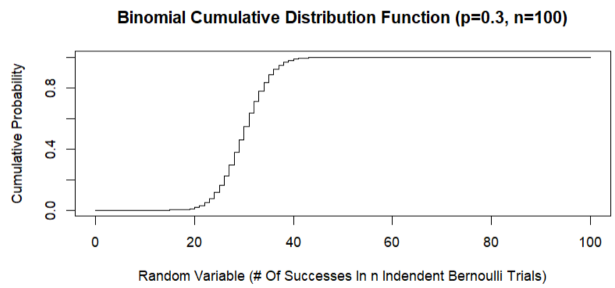

Figure 3.5: Binomial CDF Example

```
######Binomial Distribution#####
#number of successes in n Bernoulli Trials, each independent with probability p#
###Parameters###
p=0.3                              #the probability of success (e.g. 'true' batting average)#
trials=5                           #number of hits in 5 at bats#

###PMF###
x=seq(from=0, to=trials, by=1)
y=dbinom(x, size=trials, prob=p)   #use Rlab package#
plot(x, y, type="h",               #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     lwd=5, col="blue",ylab="Probability",
     xlab="Random Variable (# Of Successes In n Indepenent Bernoulli Trials)",
     main=paste0("Binomial Probability Mass Function (p=", p, ", n=", trials, ")"))

###CDF###
x=seq(from=0, to=100, by=1)
y=pbinom(x, size=100, prob=p)
plot(x, y, type="s",               #s for step, l for line#
     xlim=c(min(x)-0.1,max(x)+0.1), ylim=c(0,1),
     ylab="Cumulative Probability",
     xlab="Random Variable (# Of Successes In n Indendent Bernoulli Trials)",
     main=paste0("Binomial Cumulative Distribution Function (p=", p, ", n=100)"))

###Simulations And Questions###
mysim=rbinom(100, 100, p)         #100 random simulations from Binomial#
pbinom(30, size=100, prob=p)      #probability get 30 or less in 100 trials of prob=p#
```

Figure 3.6: Binomial R Script

3) Geometric Distribution: The random variable is the total number of failures before the first success in a succession of independent Bernoulli Trials. For example, the random variable could be the number of hitless at bats baseball player with a "true" batting average of 0.3 has before his first hit.

Notation and Parameter(s): $X \sim \text{Geom}(p)$

Support: $k \in \mathbb{N}$

PMF: $p(k) = (1-p)^{k-1}p$

CDF: $F(k) = \begin{cases} 0, & k < 0 \\ 1-(1-p)^{\lfloor k \rfloor + 1}, & k \geq 0 \end{cases}$

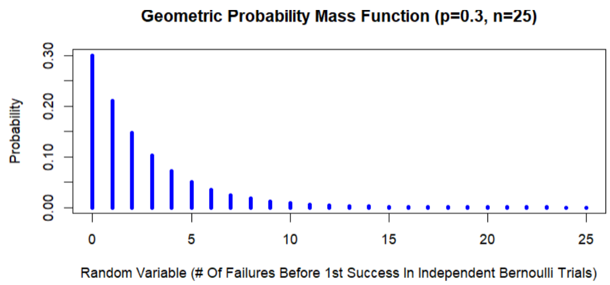

Figure 3.7: Geometric PMF Example

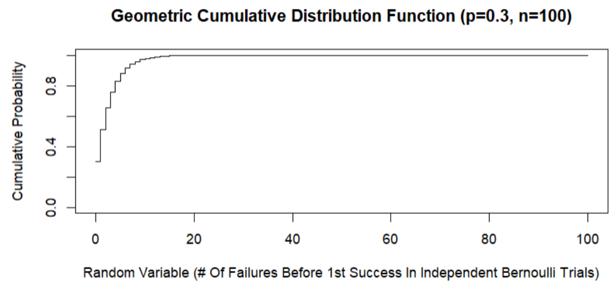

Figure 3.8: Geometric CDF Example

```
######Geometric Distribution#####
#number of independent failures with probability 1-p before 1st success#
#"shifted" geometric is when define as number of trials for 1st success#

###Parameters###
p=0.3                           #the probability of success (e.g. 'true' batting average)#
trials=25                       #e.g. number of at bats before 1st hit#

###PMF###
x=seq(from=0, to=trials, by=1)
y=dgeom(x, prob=p)              #use Rlab package#
plot(x,y,type="h",              #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=5, col="blue", ylab="Probability",
     xlab="Random Variable (# Of Failures Before 1st Success In Independent Bernoulli Trials)",
     main=paste0("Geometric Probability Mass Function (p=", p, ", n=", trials, ")"))

###CDF###
x=seq(from=0, to=100, by=1)
y=pgeom(x, prob=p)
plot(x,y,type="s",              #s for step, l for line#
     xlim=c(min(x)-0.1,max(x)+0.1), ylim=c(0,1),
     ylab="Cumulative Probability",
     xlab="Random Variable (# Of Failures Before 1st Success In Independent Bernoulli Trials)",
     main=paste0("Geometric Cumulative Distribution Function (p=", p, ", n=100)"))

###Simulations And Questions###
mysim=rgeom(100, p)            #100 random simulations from Geometric#
pgeom(30, prob=p)              #probability get 30 or less in 100 trials of prob=p#
```

Figure 3.9: Geometric R Script

4) Negative Binomial: A generalization of the geometric, the random variable is the total number of failures before k successes in a succession of independent Bernoulli Trials. For example, the random variable could be the number of at bats baseball player with a "true" batting average of 0.3 has before getting five hits.

Notation and Parameter(s): $X \sim \text{NB}(p, r)$ Support: $k \in \mathbb{N}$
PMF: $p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$

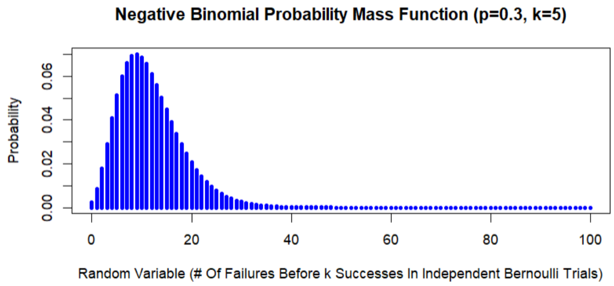

Figure 3.10: Negative Binomial PMF Ex.

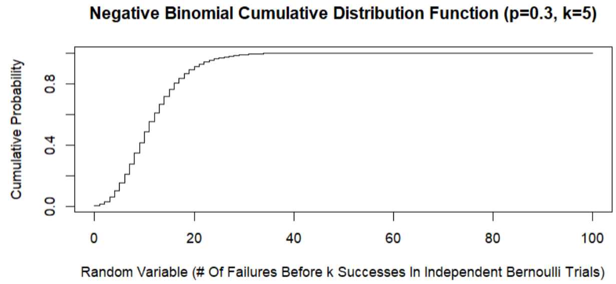

Figure 3.11: Negative Binomial CDF Ex.

```
######Negative Binomial (Pascal) Distribution#####
#generalizes geometric#
#"How many independent trials w/ prob of success p until k success?"#

###Parameters###
p=0.3                                   #the probability of success (e.g. 'true' batting average)#
successes=5                             #e.g. number of hitless at bats before 5th hit#

###PMF###
x=seq(from=0, to=100, by=1)
y=dnbinom(x, size=successes, prob=p)    #use Rlab package#
plot(x,y,type="h",                      #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=5, col="blue", ylab="Probability",
     xlab="Random Variable (# Of Failures Before k Successes In Independent Bernoulli Trials)",
     main=paste0("Negative Binomial Probability Mass Function (p=", p, ", k=", successes, ")"))

###CDF###
x=seq(from=0, to=100, by=1)
y=pnbinom(x, size=successes, prob=p)
plot(x,y,type="s",                      #s for step, l for line#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability",
     xlab="Random Variable (# Of Failures Before k Successes In Independent Bernoulli Trials)",
     main=paste0("Negative Binomial Cumulative Distribution Function (p=", p, ", k=", successes, ")"))

###Simulations And Questions###
mysim=rnbinom(n=100, size=5, prob=p)     #100 random simulations from Negative Binomial#
pnbinom(5, size=100, prob=p)             #probability get 5 or less in 100 trials of prob=p#
```

Figure 3.12: Negative Binomial R Script

5) Hypergeometric: The random variable is the total number of successes when sampling m elements without replacement from a set of size n, of which there are r desired elements. For example, the random variable could be the number of matched numbers lottery numbers after selecting 6 numbers from a set of 100.

Notation and Parameter(s): $X \sim \text{HG}(r, m, n)$ Support: $k \in \{0, 1, \ldots, r\}$

PMF: $p(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$

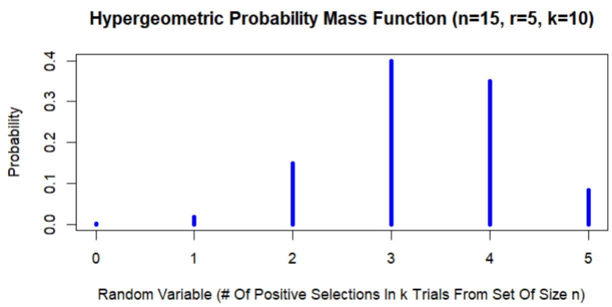

Figure 3.13: Hypergeometric PMF Example

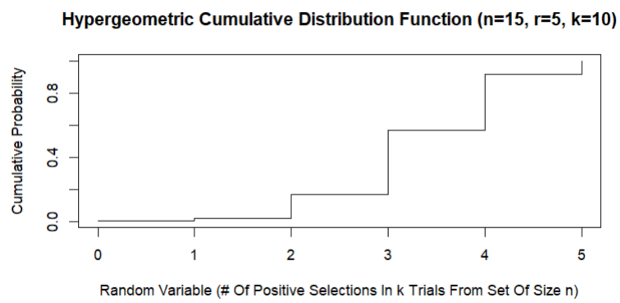

Figure 3.14: Hypergeometric CDF Example

```
######Hypergeometric Distribution#####
#"Of set with n objects, r of which are desirable,...#
#...how many of r are chosen after m selections (w/o replacement)?"#

###Parameters###
population=15                               #e.g. number of socks in a drawer#
success=5                                   #e.g. number of black socks#
selection=10                                #e.g. number of socks picked from drawer#

###PMF###
x=seq(from=0, to=success, by=1)
y=dhyper(x, m=success, n=population-success , k=selection)        #use Rlab package#
plot(x,y,type="h",                          #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=5, col="blue", ylab="Probability",
     xlab="Random Variable (# Of Positive Selections In k Trials From Set Of Size n)",
     main=paste0("Hypergeometric Probability Mass Function (n=", population,
                 ", r=", success, ", k=", selection, ")"))

###CDF###
x=seq(from=0, to=success, by=1)
y=phyper(x, m=success, n=population-success , k=selection)
plot(x,y,type="s",                         #s for step, l for line#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability",
     xlab="Random Variable (# Of Positive Selections In k Trials From Set Of Size n)",
     main=paste0("Hypergeometric Cumulative Distribution Function (n=", population,
                 ", r=", success, ", k=", selection, ")"))

###Simulations And Questions###
mysim=rhyper(100, m=success, n=population-success , k=selection)   #100 random simulations from Hypergeometric#
phyper(0, m=success, n=population-success , k=selection)           #probability get all 10 white socks=0.03%#
```

Figure 3.15: Hypergeometric R Script

6) Poisson: The random variable is the number of events occurring in a fixed interval t, should the occurrence of the events occur independently of the last event, and should the mean number of events occurring in any time period be known. For example, the random variable could be the number of calls received in a twenty minute increment at a call center that receives an average of 10 calls an hour.

Notation and Parameter(s): $X \sim \text{Pois}(\lambda)$ Support: $k \in \mathbb{N}$

PMF: $p(k) = \frac{\lambda^k}{k!}e^{-\lambda}$

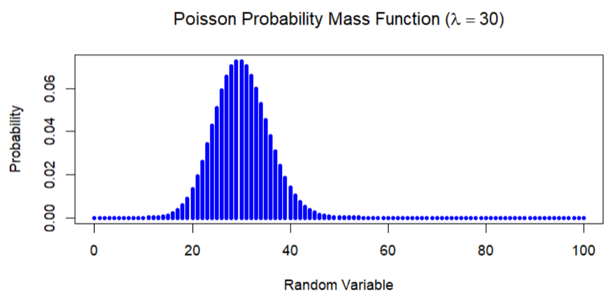

Figure 3.16: Poisson PMF Example

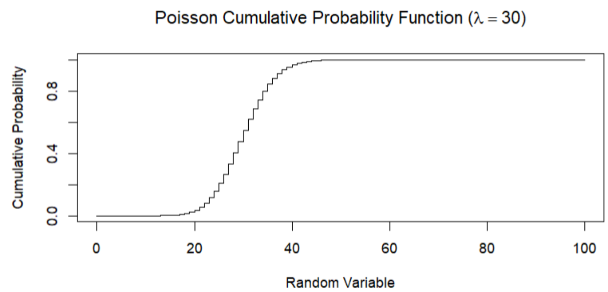

Figure 3.17: Poisson CDF Example

```
######Poisson Distribution#####
#"What is prob get x in y time given average n"#
#Limit of Binomial Distribution W/ large n and small p#

###Parameters###
mylambda=30                                    #e.g. number of calls an hour#

###PMF###
x=seq(from=0, to=100, by=1)
y=dpois(x, mylambda)                           #use Rlab package#
plot(x,y,type="h",                             #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=5, col="blue", ylab="Probability", xlab="Random Variable",
     main=bquote("Poisson Probability Mass Function (" * lambda==.(mylambda) * ")"))

###CDF###
x=seq(from=0, to=100, by=1)
y=ppois(x, mylambda)
plot(x,y,type="s",                    #s for step, l for line#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Poisson Cumulative Probability Function (" * lambda==.(mylambda) * ")"))

###Simulations And Questions###
mysim=rpois(100, mylambda)                     #100 random simulations from Poisson#
ppois(3, mylambda)                             #probability get all three calls in 1 hr span#
```

Figure 3.18: Poisson R Script

7) Discrete Uniform: The random variable is an integer from the range $[a, b]$. For example, the random variable could be the number rolled from a fair dice, and the support would be the integers one through six.

Notation and Parameter(s): $X \sim \text{DU}(a, b)$ Support: $k \in \{a, a{+}1, \ldots, b\}$
PMF: $p(k) = \frac{1}{b-a}$

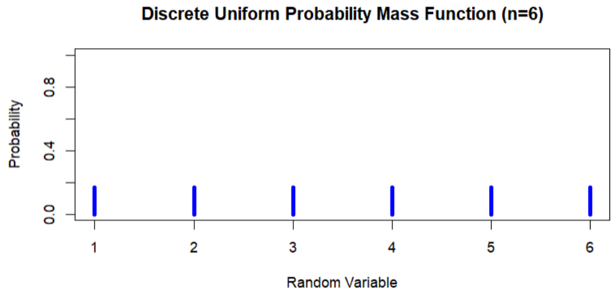

Figure 3.19: Discrete Uniform PMF Example

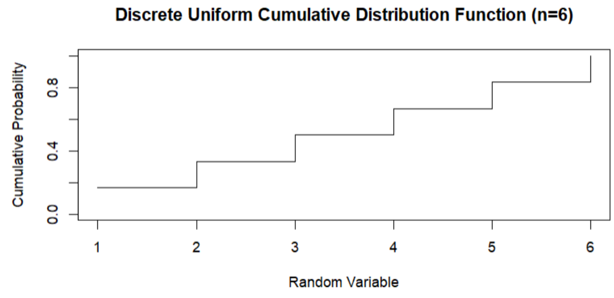

Figure 3.20: Discrete Uniform CDF Example

```
######Discrete Uniform Distribution#####
###Parameters###
size=6                                                      #fair dice#

###PMF###
x=seq(from=1, to=size, by=1)
y=rep(1/size, size)
plot(x,y,type="h",                          #h for histogram#
     xlim=c(min(x),max(x)), ylim=c(0, 1),
     lwd=5, col="blue", ylab="Probability", xlab="Random Variable",
     main=paste0("Discrete Uniform Probability Mass Function (n=", size, ")"))

###CDF###
x=seq(from=1, to=size, by=1)
y[1]=1/size; for (i in 2:size) {
  y[i]=y[i-1]+y[i]
}
plot(x,y,type="s",                          #s for step, l for line#
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=paste0("Discrete Uniform Cumulative Distribution Function (n=", size, ")"))

###Simulations And Questions###
mysim=sample(size, 100, replace=T)                #100 random simulations from uniform#
```

Figure 3.21: Discrete Uniform R Script

3.2 Continuous Random Variables

1) Continuous Uniform: The random variable is a real number between a and b. For example, the random variable could be a number between 0 and 1.

Notation and Parameter(s): $X \sim \mathrm{U}(a,b)$

Support: $x \in [a,b]$

PDF: $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & x \notin [a,b] \end{cases}$

CDF: $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$

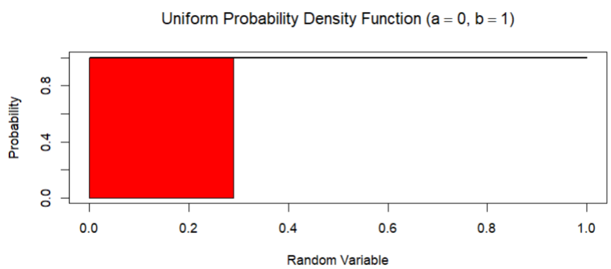

Figure 3.22: Uniform PDF Example

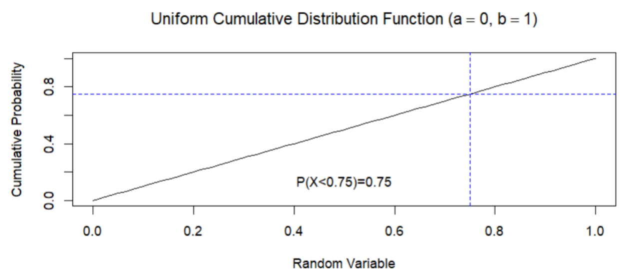

Figure 3.23: Uniform CDF Example

```
######Continuous Uniform Distribution#####
###Parameters###
a=0 ; b=1                                               #the range#

###PMF###
x=seq(from=a, to=b, by=.01) ; y=dunif(x) ; pdf=data.frame(x,y)
plot(pdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=2, col="black", ylab="Probability", xlab="Random Variable",
     main=bquote("Uniform Probability Density Function (" *
                   a==.(a) * "," ~ b==.(b) *")"))

###CDF###
x=seq(from=a, to=b, by=.01); y=punif(x); cdf=data.frame(x,y)
plot(cdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Uniform Cumulative Distribution Function (" *
                   a==.(a) * "," ~ b==.(b) *")"))

###Simulations And Questions###
val=0.3 ; myquart=0.75                                   #example values#
mysim=runif(100, min=a, max=b)                           #100 random simulations from uniform#
punif(0.3, min=a, max=b)                                 #probability less than 0.3#

x1=min(which(pdf$x >= min(x))); x2=max(which(pdf$x < val))
with(pdf, polygon(x=c(x[c(x1,x1:x2,x2)]), y= c(0, y[x1:x2], 0), col="red"))

q=qunif(myquart, min=a, max=b)                            #the 3rd quartile#
with(cdf, abline(h=myquart , lty=2, col="blue"))
with(cdf, abline(v=q , lty=2, col="blue"))
mtext(paste0("P(X<", round(q,2), ")=", myquart), side=1, line=-2)
```

Figure 3.24: Continuous Uniform R Script

2) Exponential Distribution: The random variable is the lifetime of a memory-less object (i.e. the probability the objects last t years after already lasting s years is the same as the probability the object lasts t years after already lasting $s+1$ years). For example, the random variable could be the length of time a computer lasts.

Notation and Parameter(s): $X \sim \text{Exp}(\lambda)$ Support: $x \in \mathbb{R}_+$

PDF: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ CDF: $F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0, & x < 0 \end{cases}$

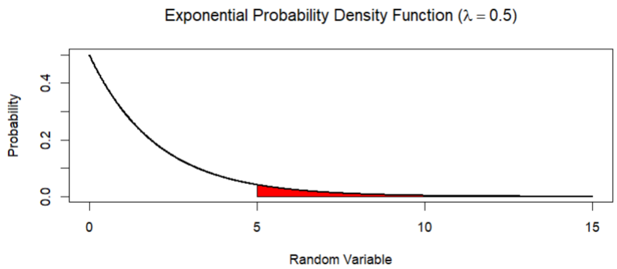

Figure 3.25: Exponential PDF Example

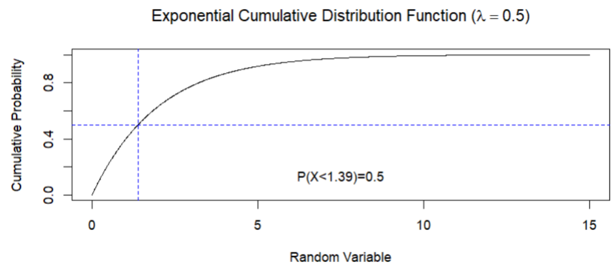

Figure 3.26: Exponential CDF Example

```
######Exponential Distribution#####
###Parameters###
mylambda=0.5                          #~n*p in binom#

###PMF###
x=seq(from=0, to=15, by=.01) ; y=dexp(x, mylambda) ; pdf=data.frame(x,y)
plot(pdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=2, col="black", ylab="Probability", xlab="Random Variable",
     main=bquote("Exponential Probability Density Function (" *
                  lambda==.(mylambda)*")"))

###CDF###
x=seq(from=0, to=15, by=.01); y=pexp(x, mylambda); cdf=data.frame(x,y)
plot(cdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Exponential Cumulative Distribution Function (" *
                  lambda==.(mylambda) *")"))

###Simulations And Questions###
val=5 ; myquart=.5                    #example values#
mysim=rexp(100, mylambda)             #100 random simulations from exponential#
1-pexp(val, mylambda)                 #probability greater than value#

x1=min(which(pdf$x >= val)); x2=max(which(pdf$x < max(x)))
with(pdf, polygon(x=c(x[c(x1,x1:x2,x2)]), y= c(0, y[x1:x2], 0), col="red"))

q=qexp(myquart, mylambda)             #the median#
with(cdf, abline(h=myquart , lty=2, col="blue"))
with(cdf, abline(v=q , lty=2, col="blue"))
mtext(paste0("P(X<", round(q,2), ")=", myquart), side=1, line=-2)
```

Figure 3.27: Exponential R Script

3) Gamma Distribution: The random variable could be the time left until death of a person.

Notation and Parameter(s): $X \sim \text{Gamma}(\alpha, \lambda)$ Support: $t \in [a, b]$
PDF: $g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$

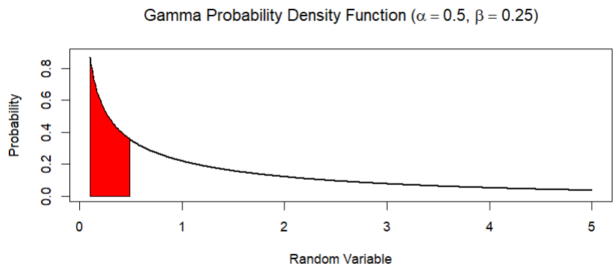

Figure 3.28: Gamma PDF Example

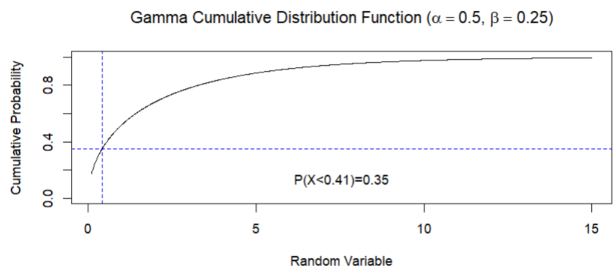

Figure 3.29: Gamma CDF Example

```
######Gamma Distribution#####
###Parameters###
shape=0.5; scale=0.25

###PMF###
x=seq(from=0.1, to=5, by=.01) ; y=dgamma(x, shape, scale) ; pdf=data.frame(x,y)
plot(pdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=2, col="black", ylab="Probability", xlab="Random Variable",
     main=bquote("Gamma Probability Density Function (" *
                      alpha==.(shape) *"," ~ beta==.(scale) * ")"))

###CDF###
x=seq(from=0.1, to=15, by=.01); y=pgamma(x, shape, scale); cdf=data.frame(x,y)
plot(cdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Gamma Cumulative Distribution Function (" *
                      alpha==.(shape) *"," ~ beta==.(scale) * ")"))

###Simulations And Questions###
val=0.5 ; myquart=.35                          #example values#
mysim=rgamma(100, shape, scale)                #100 random simulations from gamma#
pgamma(val, shape, scale)                      #probability less than value#

x1=min(which(pdf$x >= min(x))); x2=max(which(pdf$x < val))
with(pdf, polygon(x=c(x[c(x1,x1:x2,x2)]), y= c(0, y[x1:x2], 0), col="red"))

q=qgamma(myquart, shape, scale)                 #the 35th quantile#
with(cdf, abline(h=myquart , lty=2, col="blue"))
with(cdf, abline(v=q , lty=2, col="blue"))
mtext(paste0("P(X<", round(q,2), ")=", myquart), side=1, line=-2)
```

Figure 3.30: Gamma R Script

4) Beta Distribution: The random variable is essentially a distribution on probability itself. For example, the random variable could be the probability that the probability someone watches an add is over 50%.

Notation and Parameter(s): $X \sim \text{Beta}(a, b)$ Support: $x \in [0, 1]$

PDF: $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$

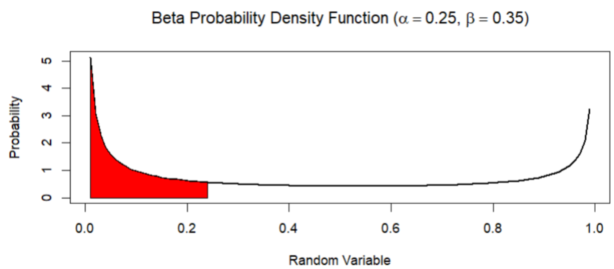

Figure 3.31: Beta PDF Example

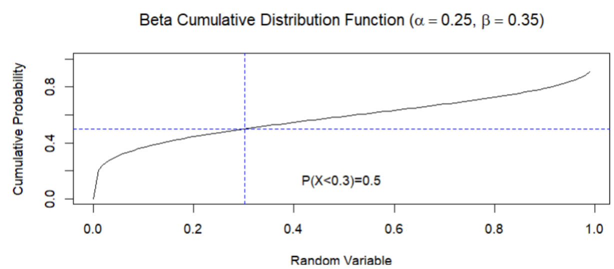

Figure 3.32: Beta CDF Example

```
######Beta Distribution#####
###Parameters###
shape=0.25; scale=0.35

###PMF###
x=seq(from=0.01, to=.99, by=.01) ; y=dbeta(x, shape, scale) ; pdf=data.frame(x,y)
plot(pdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=2, col="black", ylab="Probability", xlab="Random Variable",
     main=bquote("Beta Probability Density Function (" *
                  alpha==.(shape) *"," ~ beta==.(scale) * ")"))

###CDF###
x=seq(from=0.01, to=.99, by=.01); y=pbeta(x, shape, scale); cdf=data.frame(x,y)
plot(cdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Beta Cumulative Distribution Function (" *
                  alpha==.(shape) *"," ~ beta==.(scale) * ")"))

###Simulations And Questions###
val=0.25 ; myquart=.5                        #example values#
mysim=rbeta(100, shape, scale)               #100 random simulations from beta#
pbeta(val, shape, scale)                     #probability less than value#

x1=min(which(pdf$x >= min(x))); x2=max(which(pdf$x < val))
with(pdf, polygon(x=c(x[c(x1,x1:x2,x2)]), y= c(0, y[x1:x2], 0), col="red"))

q=qbeta(myquart, shape, scale)               #the median#
with(cdf, abline(h=myquart , lty=2, col="blue"))
with(cdf, abline(v=q , lty=2, col="blue"))
mtext(paste0("P(X<", round(q,2), ")=", myquart), side=1, line=-2)
```

Figure 3.33: Beta R Script

5) Normal Distribution: The random variable could be the height of a person.

Notation and Parameter(s): $X \sim \text{N}(\mu, \sigma^2)$ Support: $x \in \mathbb{R}$

PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

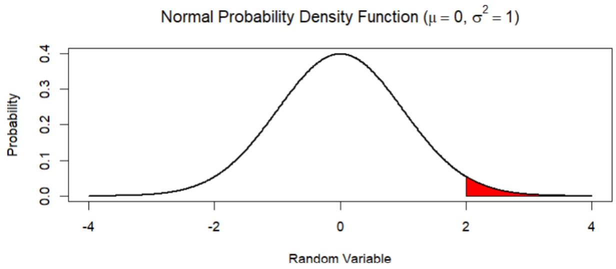

Figure 3.34: Normal PDF Example

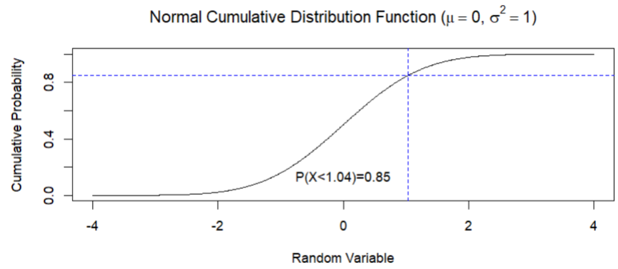

Figure 3.35: Normal CDF Example

```
######Normal Distribution#####
###Parameters###
mymean=0; myvar=1

###PMF###
x=seq(from=-4, to=4, by=.01) ; y=dnorm(x, mymean, myvar) ; pdf=data.frame(x,y)
plot(pdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0, max(y)),
     lwd=2, col="black", ylab="Probability", xlab="Random Variable",
     main=bquote("Normal Probability Density Function (" *
                   mu==.(mymean) *"," ~ sigma^2==.(myvar) * ")"))

###CDF###
x=seq(from=-4, to=4, by=.01) ; y=pnorm(x, mymean, myvar); cdf=data.frame(x,y)
plot(cdf, type="l",
     xlim=c(min(x),max(x)), ylim=c(0,1),
     ylab="Cumulative Probability", xlab="Random Variable",
     main=bquote("Normal Cumulative Distribution Function (" *
                   mu==.(mymean) *"," ~ sigma^2==.(myvar) * ")"))

###Simulations And Questions###
val=2 ; myquart=.85                              #example values#
mysim=rnorm(100, mymean, myvar)                  #100 random simulations from normal#
1-pnorm(val, mymean, myvar)                      #probability greater than value#

x1=min(which(pdf$x >= val)); x2=max(which(pdf$x < max(x)))
with(pdf, polygon(x=c(x[c(x1,x1:x2,x2)]), y= c(0, y[x1:x2], 0), col="red"))

q=qnorm(myquart, mymean, myvar)                   #the 85th quantile#
with(cdf, abline(h=myquart , lty=2, col="blue"))
with(cdf, abline(v=q , lty=2, col="blue"))
mtext(paste0("P(X<", round(q,2), ")=", myquart), side=1, line=-2)
```

Figure 3.36: Normal R Script

4 Transformations And Approximations

It may be the case that one knows a random variable X, but wants to study values of interest related to X. For example, if one knows the distribution for the amount of people who enter a store every hour, they may be interested in understanding the revenue the store gets every hour, which they believe to be a function of X. In general, methods for finding distributions of transformations can be time-consuming, so if only certain properties of the transformation are of interest (such as mean and variance), it may be worthwhile to estimate these values instead.

4.1 Theorems And Further Examples

Theorem 4.1. CDF Method: Given a random variable X and a transformation $Y = g(X)$, we can calculate the CDF of Y as $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$.

Theorem 4.2. MGF Method: Given a random variable X and a transformation $Y = g(X)$, we can calculate the MGF of Y as $M_Y(y) = \mathbb{E}\left(e^{tY}\right) = \mathbb{E}\left(e^{tg(X)}\right)$, and then use properties of exponents and expectations to break apart the expected value into (hopefully) the form of some known MGF.

Theorem 4.3. Jacobian Method: Given a continuous random variable X and a differentiable, monotonically increasing transformation $Y = g(X)$, we can calculate the PDF of Y as $f_Y(y) = f_x\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|$.

Theorem 4.4. Probability Integral Transform: Given a continuous random variable X, the transformation $Y = F_X(X)$ yields a standard $U(0, 1)$ distribution.

Proof. $F_Y(y) = P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$ ■

Theorem 4.5. Generate Random Variables: Given a continuous random variable X and a source for psuedo-random numbers from a standard uniform distribution U, we can generate random values from the distribution X by computing $Y = F_X^{-1}(U)$ and realizing that the CDF of Y is F_X.

Proof. $F_Y(y) = P(Y \leq y) = P(F_X^{-1}(U) \leq y) = P(U \leq F_X(y)) = F_U(F_X(y)) = F_X(y)$ ■

Theorem 4.6. Propagation of Error: Given a random variable X with known mean and variance, and given the transformation $Y{=}g(X)$, we can get 1^{st} order approximations for the mean and variance of Y with the formula $\mathbb{E}(Y){\approx}g\left(\mathbb{E}(X)\right)$ and $\mathbb{V}(Y) \approx \mathbb{V}(X)\left[g'\left(\mathbb{E}(X)\right)\right]^2$. We can get improved 2^{nd} order approximations with the formula $\mathbb{E}(Y) \approx g\left(\mathbb{E}(X)\right) + \frac{1}{2}\mathbb{V}(X)g''\left(\mathbb{E}(X)\right)$.

Proof. Use Taylor Expansion about $\mathbb{E}(X)$, group terms, and then apply properties of expectations on linear functions (e.g. $Y{=}aX{+}b \Longrightarrow \mathbb{E}(Y){=}a\mathbb{E}(X){+}b$, see Theorems 2.1 and 2.2).

$$g(X)=\sum_{n=0}^{\infty}\frac{g^{(n)}(\mu_X)}{n!}(X-\mu_X)^n \implies g(X) \approx g(\mu_X)+g'(\mu_X)(X-\mu_X)=\left[g'(\mu_X)\right]X+\left[g(\mu_X)-g'(\mu_X)\mu_X\right]$$

■

4.2 Problems

4.30) Find $\mathbb{E}[1/(X+1)]$ where X is a Poisson random variable.

The probability mass function of a Poisson random variable X is $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$. Using Taylor's Series for e^x, we know $e^x = \sum\limits_{n=0}^{\infty} \frac{x^n}{n!}$ and so calculate the expected value of X as

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=1}^{\infty} k \cdot \frac{\lambda\lambda^{k-1} e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{k\lambda^{k-1}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{(k-1)}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Then as a function of the random variable, the expected value is calculated as

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{X+1}\right) &= \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{(k+1)!} e^{-\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} \frac{e^{-\lambda}}{\lambda} \\
&= \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}}{\lambda} \left(\sum_{y=1}^{\infty} \frac{\lambda^y}{y!} \right) \\
&= \frac{e^{-\lambda}}{\lambda} \left(-1 + \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} \right) = \frac{e^{-\lambda}}{\lambda} \left(-1 + e^{\lambda} \right) = \frac{1 - e^{-\lambda}}{\lambda}
\end{aligned}$$

2.56) If $X \sim N(0, \sigma^2)$, find the density of $Y = |X|$

The Cumulative Distribution Function (CDF) of Y is $F_Y(y) = P(Y \leq y) = P(|X| \leq y)$. By the properties of absolute value, this is $P(-y \leq X \leq y) = P(X \leq y) - P(X \leq -y)$. Each term is in the form of a CDF for X, so we can write $F_Y(y) = F_X(y) - F_X(-y)$.

Recall $f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y)$ and that the Probability Density Function (PDF) of a normal distribution with mean μ and standard deviation σ is $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. So we have:

$$\begin{aligned}
f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\big(F_X(y) - F_X(-y)\big) && \text{substituting from above} \\
&= f_X(y)\left(\frac{\mathrm{d}}{\mathrm{d}y} y\right) - f_X(-y)\left(\frac{\mathrm{d}}{\mathrm{d}y} -y\right) && \text{chain rule} \\
&= f_X(y) + f_X(-y) && \text{simplifying} \\
&= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{y}{\sigma}\right)^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{-y}{\sigma}\right)^2} && X \text{ is normal with mean } 0 \\
&= \frac{2}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}(\frac{y}{\sigma})^2}, y \in [0, \infty) && \text{and zero otherwise}
\end{aligned}$$

2.66) Let $f_X(x) = \alpha x^{-(\alpha+1)}$ for $x \geq 1$ and $f_X(x) = 0$ otherwise, where α is a positive parameter. Show how to generate random variables from this density from a uniform random number generator.

We first find the CDF of the distribution:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{1}^{x} \alpha t^{-(\alpha+1)}\, dt = -t^{-\alpha}\,\Big|_1^x = 1 - \frac{1}{x^\alpha}$$

We next calculate the inverse:

$$\left[F_X(x) = 1 - \frac{1}{x^\alpha}\right] \implies \left[x = 1 - \frac{1}{\left(F_X^{-1}(x)\right)^\alpha}\right] \implies F_X^{-1}(x) = \left(\frac{1}{1-x}\right)^{1/\alpha}$$

Finally, we take samples u_i from a continuous uniform distribution $U \sim U(0,1)$ and subsequently evaluate F_X^{-1} at each sample point. The resultant values $F_X^{-1}(u_i)$ are random variables from the density f_X. For example:

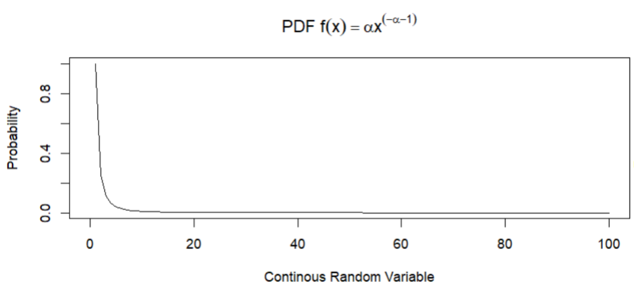

Figure 4.1: PDF

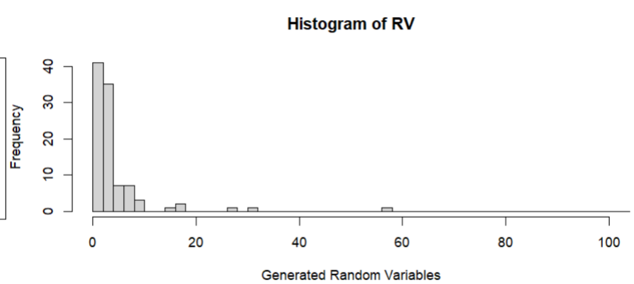

Figure 4.2: Generated Random Variables

```
a=1                                     #example parameter#
n=100                                   #sample size#
pdf=function(x,y) {y*(x^(-y-1))}        #pdf given#
pdfx=seq(from=1, to=n)
pdfy=pdf(pdfx,a)
plot(pdfx, pdfy, type="l",
     xlim=c(0,max(pdfx)), ylim=c(min(pdfy), max(pdfy)),
     xlab="Continous Random Variable", ylab="Probability",
     main=expression(plain("PDF ") *f(x)==alpha * x^(-alpha-1)))

set.seed(501)                           #so can precisely replicate#
u=runif(n=100, min=0, max=1)            #n random samples from (0,1) uniform dist#
f.inv=function(x,y) {((1)/(1-x))^(1/y)} #inverse of cdf#
invy=f.inv(u,a)
hist(invy, xlim=c(0,max(pdfx)), breaks=100,
     xlab="Generated Random Variables",
     main="Histogram of RV")
```

Figure 4.3: R Script

4.100) If X is a uniform on $[10, 20]$, find the approximate and exact mean and variance of $Y = 1/X$ and compare them.

Given the 1st and 2nd central moments of a distribution X, call them μ_X and σ_X^2, we are interested in the mean and variance of a transformation $Y = g(X)$. When the transformation is linear, i.e. can be written $Y = aX + b$ for some real a, b, then by properties of expectations:

$$\mu_Y = \mathbb{E}(Y) = \mathbb{E}(aX + b) = a\mathbb{E}(X) + b = a\mu_X + b \quad (4.1)$$
$$\sigma_Y^2 = \mathbb{V}(Y) = \mathbb{V}(aX + b) = a^2\mathbb{V}(X) = a^2\sigma_X^2 \quad (4.2)$$

The Taylor Series approximation for a generic transformation g about the mean μ_X is given by $g(X) = \sum_{n=0}^{\infty} \frac{g^{(n)}(\mu_X)}{n!}(X - \mu_X)^n$. This gives a linear approximation for $Y = g(x)$. Up to the first order, the approximation is:

$$g(X) \approx g(\mu_X) + g'(\mu_X)(X - \mu_X) = [g'(\mu_X)]\, X + [g(\mu_X) - g'(\mu_X)\mu_X]$$

Then from equation 4.1 and Equation 4.2 above:

$$\mu_Y \approx [g'(\mu_X)]\, \mathbb{E}(X) + [g(\mu_X) - g'(\mu_X)\mu_X] = g'(\mu_X)\mu_X + g(\mu_X) - g'(\mu_X)\mu_X = g(\mu_X)$$
$$\sigma_Y \approx [g'(\mu_X)]^2\, \sigma_X^2$$

Proceeding with our approximation, we first calculate the mean and variance of X.

$$\mathbb{E}(X) = \frac{1}{2}(20 + 10) = 15$$
$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \left(\int_{10}^{20} x^2 \cdot \frac{1}{20 - 10}\, dx\right) - 15^2$$
$$= \left(\frac{1}{10} \cdot \frac{x^3}{3} |_{10}^{20}\right) - 225 = \frac{20^3 - 10^3}{30} - 225 = \frac{7000 - 6750}{30} = \frac{25}{3}$$

So our approximations are $\mu_Y \approx \frac{1}{\mu_X} = \frac{1}{15}$ and $\sigma_Y^2 \approx \left(\frac{-1}{15^2}\right)^2 \frac{25}{3} = \frac{25}{3 \cdot 15^4} = \frac{1}{6075}$.

To get the exact values, we use the CDF method to find the distribution and calculate moments.

$$F_Y(y) = P(Y \le y) = P\Big(\frac{1}{X} \le y\Big) \qquad y \text{ positive since } x \text{ positive}$$
$$= P\Big(X \ge \frac{1}{y}\Big) = 1 - P\Big(X \le \frac{1}{y}\Big) = 1 - F_X\Big(\frac{1}{y}\Big) \qquad \text{The CDF}$$
$$f_Y(y) = \frac{d}{dy}(F_Y(y)) = -f_X\Big(\frac{1}{y}\Big) \cdot \frac{-1}{y^2} = \frac{1}{y^2} f_X\Big(\frac{1}{y}\Big) = \frac{1}{y^2(20 - 10)} \qquad \text{The PDF}$$
$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \frac{1}{10y^2}\, dy = \frac{1}{10}\int_{\frac{1}{20}}^{\frac{1}{10}} \frac{1}{y}\, dy = \frac{1}{10}\left[\ln(y) - \ln(y)|_{.05}^{.1}\right] = \frac{\ln(2)}{10} \qquad \text{Exact mean}$$
$$\mathbb{V}(Y) = \Big(\int_{1/20}^{1/10} y^2 \frac{1}{10y^2}\Big) - \frac{\ln(2)^2}{100} = \frac{1}{100} - \frac{1}{200} - \frac{\ln(2)^2}{100} = \frac{1 - 2\ln(2)^2}{200} \qquad \text{Exact Variance}$$

1) Let Y have PDF $f_Y(y, \theta) = \frac{1}{\theta} y^{1/\theta - 1}$ for $0 < y < 1$ and $\theta > 0$. Establish the distribution of $X = -\ln(Y)$.

It is clear that f is a continuous random variable. Just as clearly, the negative natural log transformation $g(y) = -\ln(y)$ is differentiable and monotonically decreasing on $\mathbb{R}_+ \backslash \{0\}$ (and so on the support of y). These conditions fulfill the criterion for the Jacobian Method, and so the distribution of X is given by $f_X(x) = f_Y\left(g^{-1}(x)\right) \left|\frac{d}{dx} g^{-1}(x)\right|$.

The inverse is calculated as $g(x) = -\ln(x) \implies x = -\ln(g^{-1}(x)) \implies g^{-1}(x) = e^{-x}$ and so the derivative of the inverse is $-e^{-x}$.

Plugging these calculations into the above formula, we have $f_X(x) = f_Y(e^{-x}) \cdot e^{-x}$. Finally we see $f_X(x) = \frac{1}{\theta} e^{-x^{\frac{1}{\theta} - 1}} \cdot e^{-x} = \frac{1}{\theta} \cdot e^{-(\frac{1}{\theta})x}$. This is the exponential distribution with parameter $\lambda = \frac{1}{\theta}$; $X \sim \mathrm{Exp}\left(\frac{1}{\theta}\right)$. Since the support of Y is $(0, 1)$, and since the natural log of y goes to $-\infty$ as y approaches 0 from the right, the support of X is $(0, \infty)$.

5 Joint Distributions

5.1 Definitions

Definition 5.1. Joint Distributions: Often, one is interested in studying random variables which have the same probability space. For example, one may be interested in random variables for adult height, and the average of the adult's parent's height. In general, the properties and rules for the multivariate case are largely identical to the univariate case. For example, the **joint density function** $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy$ integrates to one and the **joint CDF** is $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$. One should take care to interpret these probabilities correctly in order to get the right surface to integrate over. For example, $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)$. All properties are identical for the discrete class just replacing integration with summation

Definition 5.2. Marginal Distributions: Given a joint distribution, it may be of interest to break the distribution into the univariate distributions. For example, $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$. Note that in contrast, one can not generally construct the joint distribution from marginal distributions. This definition becomes more clear in the discrete case. The marginal distribution is the "margin" after summing the X random variable along each Y variable.

Definition 5.3. Conditional Distributions: In addition to breaking the joint distribution into univariate distributions, one may be interested in studying the distribution of one variable conditional on another. For example $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_y(y)}$. Again this definition becomes more clear in the discrete case. We condition on the y desired, then divide each entry along the x cases by the total y case at the given y.

Definition 5.4. Conditional Expectations: As would be expected, $\mathbb{E}(Y \mid X) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\,dy$.

Definition 5.5. Covariance: A measure of joint variability; the multivariate analog to variance in the univariate case. $\text{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Definition 5.6. Correlation ($\boldsymbol{\rho}$): A dimensionless measure of association between -1 and 1. $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$.

Definition 5.7. Independent And Identically Distributed (iid): Random Variables are said to be independent and identically distributed if the random variables are both independent and share the same distribution. For example, sampling without replacement is an identically distributed process but clearly not an independent one. However sampling with replacement is an iid process.

Definition 5.8. Order Statistics: We may be interested in the order of a collection of random variables. For example, if X_i are all iid random variables, an question of interest might be the distribution of the minimum of the X_i's. Finding minimums/maximums (or related "orders") comes down to following similar arguments as this: if U is a minimum, then the CDF of U is $P(U \leq u)$, and since U is the *minimum*, $U \leq u \implies u \leq X_i$. Then the CDF can be written as the product of CDFs for the iid random variables. In general, the K^{th} order statistic, denoted $X_{(k)}$ is $f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x) \left[1 - F(x)\right]^{n-k}$.

5.2 Theorems And Further Examples

Theorem 5.1. Properties of Independence: If X and Y are independent, then the joint PDFs factor into the marginal PDFs; $f_{X,Y} = f_X(x)f_Y(y)$. As a consequence, the joint CDFs must also factor into the marginal CDFs, the joint MGFs must factor into the marginal MGFs, and $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Theorem 5.2. Independence Implies Zero Covariance: If two random variables are independent, then they have zero covariance (and therefore zero correlation). The converse of this statement is untrue; random variables with zero covariance are not necessarily independent.

Proof. Let X and Y be independent. We are interested in computing $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. From Theorem 5.1, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, so this is immediately seen to be zero. ■

Theorem 5.3. Covariance Of Linear Transformations: If $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$, then $\text{Cov}(U, V) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \text{Cov}(X_i, Y_j)$. In particular, since $\text{Cov}(X, X) = \mathbb{V}(X)$, $\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y)$ and $\mathbb{V}(aX, bY) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\text{Cov}(X, Y)$

Corollary 5.3.1. If each $X_1, X_2, \ldots, X_n$ is independent, then $\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i)$

Theorem 5.4. Law of Iterated/Total Expectations: The expected value of Y can be found by summing or integrating the conditional expectations; $\mathbb{E}(Y) = \mathbb{E}\left[\mathbb{E}(Y \mid X)\right]$. The variance can be found with $\mathbb{V}(Y) = \mathbb{V}(\mathbb{E}(Y \mid X)) + \mathbb{E}(\mathbb{V}(Y \mid X))$

5.3 Problems

3.8) Let X and Y have the joint density $f(x,y) = \frac{6}{7}(x+y)^2$ for $0 \le x, y \le 1$. By integrating over appropriate regions, find $P(X > Y)$, $P(X+Y \le 1)$, and $P(X \le \frac{1}{2})$. Next find the marginal densities of X and Y as well as the two conditional densities.

We explain the first bounds (the case $P(X > Y)$) and follow the same procedure for the other two cases. The support of the joint density is over the region $(0,1) \times (0,1)$. Choosing to integrate over y first, the bound is minimally 0 and maximally the line $y = x$ (this is the condition that $X > Y$). After integrating over y, we can integrate over x where values can range from 0 to 1. So our double integral is $\int_0^1 \int_0^x f(x,y)\,dy\,dx$.

$$\begin{aligned}
P(X > Y) &= \int_0^1 \int_0^x f(x,y)\,dy\,dx = \frac{6}{7}\int_0^1 \int_0^x x^2 + 2xy + y^2\,dy\,dx \\
&= \frac{6}{7}\int_0^1 \left[x^2 y + xy^2 + \frac{1}{3}y^3 \Big|_0^x \right] dx = \frac{6}{7}\int_0^1 \frac{7}{3}x^3\,dx = \frac{6}{7}\frac{7}{3}\left[\frac{1}{4}x^4 \Big|_0^1\right] = \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
P(X + Y \le 1) &= \int_0^1 \int_0^{1-x} f(x,y)\,dy\,dx = \frac{6}{7}\int_0^1 \int_0^{1-x} x^2 + 2xy + y^2\,dy\,dx \\
&= \frac{6}{7}\int_0^1 \left[x^2 y + xy^2 + \frac{1}{3}y^3 \Big|_0^{1-x} \right] dx = \frac{6}{7}\int_0^1 \frac{1}{3} - \frac{x^3}{3}\,dx \\
&= \frac{6}{7}\left[\frac{x}{3} - \frac{x^4}{12}\Big|_0^1\right] = \frac{6}{7}\left(\frac{1}{4}\right) = \frac{6}{28} = \frac{3}{14}
\end{aligned}$$

$$\begin{aligned}
P(X \le 1/2) &= \int_0^{1/2} \int_0^1 f(x,y)\,dy\,dx = \frac{6}{7}\int_0^{1/2} \int_0^1 x^2 + 2xy + y^2\,dy\,dx \\
&= \frac{6}{7}\int_0^{1/2} \left[x^2 y + xy^2 + \frac{1}{3}y^3 \Big|_0^1 \right] dx = \frac{6}{7}\int_0^{1/2} x^2 + x + \frac{1}{3}\,dx \\
&= \frac{6}{7}\left[\frac{x^3}{3} + \frac{x^2}{2} + \frac{x}{3}\Big|_0^{1/2}\right] = \frac{6}{7}\left(\frac{1}{24} + \frac{1}{8} + \frac{1}{6}\right) = \frac{6}{7}\frac{8}{24} = \frac{6}{21} = \frac{2}{7}
\end{aligned}$$

To get the marginal densities, we integrate the joint density over the opposite variable.

$$\begin{aligned}
f_X(x,y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^1 \frac{6}{7}(x+y)^2\,dy \\
&= \frac{6}{7}\int_0^1 x^2 + 2xy + y^2\,dy = \frac{6}{7}\left[x^2 y + xy^2 + \frac{y^3}{3}\Big|_0^1\right] = \frac{6}{7}\left(x^2 + x + \frac{1}{3}\right)
\end{aligned}$$

$$\begin{aligned}
f_Y(x,y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_0^1 \frac{6}{7}(x+y)^2\,dx \\
&= \frac{6}{7}\int_0^1 x^2 + 2xy + y^2\,dx = \frac{6}{7}\left[\frac{x^3}{3} + x^2 y + xy^2\Big|_0^1\right] = \frac{6}{7}\left(\frac{1}{3} + y + y^2\right)
\end{aligned}$$

Finally, conditional densities are calculated as the joint density divided by the marginal density. So:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{6}{7}(x+y)^2}{\frac{6}{7}(\frac{1}{3}+y+y^2)} = \frac{x^2+2xy+y^2}{\frac{1}{3}+y+y^2}$$
$$f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)} = \frac{\frac{6}{7}(x+y)^2}{\frac{6}{7}(x^2+x+\frac{1}{3})} = \frac{x^2+2xy+y^2}{x^2+x+\frac{1}{3}}$$

3.14) Suppose that $f_{X,Y}(x,y) = xe^{-x(y+1)}$ for $x, y \in \mathbb{R}_+$. Find the marginal and conditional densities of X and Y. Are X and Y independent?

Since x and y are both positive, so too is $f_{X,Y}(x,y)$. Then the marginal density of X is $f_X(x,y) = F'_X(x,y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{\infty} f_{X,Y}(x,y)\,dy$ (the marginal density of Y is in the same format). We calculate the integrals below (note the the marginal density of X is an exponential distribution with $\lambda = 1$):

$$f_X(x,y) = \int_0^{\infty} f_{X,Y}(x,y)\,dy = \int_0^{\infty} xe^{-xy-x}\,dy \qquad u = -xy-x, du = -x\,dy, dy = \frac{-1}{x}du$$
$$\lim_{a\to\infty}\int_0^a xe^u\frac{-1}{x}\,du = -\lim_{a\to\infty}\int_0^a e^u\,du = -\lim_{a\to\infty}\left[e^{-xy-x}\Big|_0^a\right] = -\lim_{a\to\infty}\left[e^{-ax-x} - e^{-x}\right] = e^{-x}$$

$$f_Y(x,y) = \int_0^{\infty} xe^{-xy-x}\,dx \qquad u = -xy-x, du = -y\,dx - dx, dx = \frac{-1}{y+1}du$$
$$\lim_{a\to\infty}\int_0^a xe^u\frac{-1}{y+1}\,du = -\lim_{a\to\infty}\int_0^a e^u\,du = -\lim_{a\to\infty}\left[e^{-xy-x}\Big|_0^a\right] = -\lim_{a\to\infty}\left[e^{-ax-x} - e^{-x}\right] = \frac{1}{(y+1)^2}$$

Like above, we calculate the conditional densities by dividing the joint density by the marginal density and recalling exponent rules.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{xe^{-x(y+1)}}{\frac{1}{(y+1)^2}} = x(y+1)^2e^{-x(y+1)}$$
$$f_{Y|X}(y|x) = \frac{f_{Y,X}(y,x)}{f_X(x)} = \frac{xe^{-x(y+1)}}{e^{-x}} = xe^{-xy}$$

The marginal densities don't multiply to the joint density, so we know X and Y are not independent.

2) Suppose that you roll two fair six-sided dice. Let Y_1 be the number of ones and Y_2 be the number of twos. What is the join distribution of Y_1 and Y_2? What is the conditional distribution of Y_2 given $Y_1 = 0$?

The joint distribution is given below. The argument is just based on counting. There are $6^2 = 36$ possible outcomes of the roll. The number of ways that (say) there is exactly one 1 rolled and zero 2's rolled is the six possibilities for rolling 1 on the first go, less two for ensuring neither a 2 nor a 1 comes up on the second roll, plus four ways a 1 is rolled on the second roll without a 1 or 2 being rolled on the first roll; a total of eight possibilities.

		Y_1 = 0	Y_1 = 1	Y_1 = 2	
Y_2	0	16/36	8/36	1/36	25/36
	1	8/36	2/36	0	10/36
	2	1/36	0	0	1/36
		25/36	10/36	1/36	1

The conditional distribution of Y_2 given $Y_1 = 0$ is the joint probabilities from the table above divided by the marginal probability that $Y_1 = 0$; $f_{(Y_2|Y_1=0)}(y_2) = \begin{cases} \frac{16}{25} = \frac{4}{9}\frac{36}{25}, & y_2 = 0 \\ \frac{8}{25} = \frac{2}{9}\frac{36}{25}, & y_2 = 1 \\ \frac{1}{25} = \frac{1}{36}\frac{36}{25}, & y_2 = 2 \\ 0, & \text{else} \end{cases}$

4.57) If X and Y are independent random variables, find $\mathbb{V}(XY)$ in terms of the means and variance of X and Y.

$$\begin{aligned}
\mathbb{V}(XY) &= \mathbb{E}(X^2Y^2) - \mathbb{E}(XY)^2 \\
&= \mathbb{E}(X^2Y^2) - [\mathbb{E}(XY)\mathbb{E}(XY)] \\
&= \mathbb{E}(X^2)\mathbb{E}(Y^2) - [\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(X)\mathbb{E}(Y)] && \text{By independence} \\
&= \mathbb{E}(X^2)\mathbb{E}(Y^2) - \left[\mathbb{E}(X)^2\mathbb{E}(Y)^2\right] \\
&= \left[\mathbb{V}(X) + \mathbb{E}(X)^2\right]\left[\mathbb{V}(Y) + \mathbb{E}(Y)^2\right] - \left[\mathbb{E}(X)^2\mathbb{E}(Y)^2\right] && \text{Substitute for variance} \\
&= \mathbb{V}(X)\mathbb{V}(Y) + \mathbb{V}(X)\mathbb{E}(Y)^2 + \mathbb{V}(Y)\mathbb{E}(X)^2 + \mathbb{E}(X)^2\mathbb{E}(Y)^2 - \left[\mathbb{E}(X)^2\mathbb{E}(Y)^2\right] \\
&= \mathbb{V}(X)\mathbb{V}(Y) + \mathbb{V}(X)\mathbb{E}(Y)^2 + \mathbb{V}(Y)\mathbb{E}(X)^2
\end{aligned}$$

3.66) Each component of a system with three lines and two widgets has an independent exponentially distributed lifetime with parameter $\boldsymbol{\lambda}$ (i.e., the system works if at least one of the three lines has both widgets working). Find the cdf and the density of the system's lifetime.

We approach this in two parts. First, we find the distribution of the lifetime of one line in the system. The lifetime of a line is equivalent to the minimum lifetime of two widgets (minimum since a line fails if one widget fails). Next, we find the distribution for the lifetime of the system given the lifetime of a line. The lifetime of the system is equivalent to the maximum of the three lines (maximum since the system only fails when all three lines fail).

First we address the lifetime of one line. Let V be a random variable denoting the minimum lifetime of two widgets, label them X_1 and X_2. By definition, the CDF of V is $F_V(v) = P(V \leq v)$ or equivalently $F_V(v) = 1 - P(V \geq v)$. Note that $V \geq v$ if and only if $X_1, X_2 \geq v$, so $F_V(v) = 1 - [P(X_1 \geq v)P(X_2 \geq v)] = 1 - \big[\big(1 - P(X_1 \leq v)\big)\big(1 - P(X_2 \leq v)\big)\big]$. Both X_1 and X_2 have the same distribution, so this can be re-written $F_V(v) = 1 - [1 - F_X(v)]^2$

Next we address the lifetime of the system. Let U be a random variable denoting the maximum lifetime of three lines, label them V_1, V_2, V_3. By definition, the CDF of U is $F_U(u) = P(U \leq u)$ We use the same argument as above that $U \leq u$ if and only if $V_1, V_2, V_3 \leq u$. So $F_U(u) = P(V_1 \leq u)P(V_2 \leq u)P(V_2 \leq u)$ and since each of V_1, V_2, V_3 follow the same distribution, $F_U(u) = [P(V \leq u)]^3 = [F_V(u)]^3$

Since each widget $X_i \sim \text{Exp}(\lambda)$, the CDF for any one widget is $F_X(x) = 1 - e^{-\lambda x}$. Plugging this into the above derivation, we can calculate the CDF of the entire system:

$$
\begin{aligned}
F_U(u) &= F_V(u)^3 = \big[1 - [1 - F_X(u)]^2\big]^3 \\
&= \Big[1 - \big[1 - \big(1 - e^{-\lambda u}\big)\big]^2\Big]^3 = \Big[1 - \big[e^{-\lambda u}\big]^2\Big]^3 \\
&= \big[1 - e^{-2\lambda u}\big]^3 = 1 - 3e^{-2\lambda u} + 3\big(e^{-2\lambda u}\big)^2 - \big(e^{-2\lambda u}\big)^3 \\
&= 1 - 3e^{-2\lambda u} + 3e^{-4\lambda u} - e^{-6\lambda u}
\end{aligned}
$$

And then go on to calculate the density:

$$
f_U(u) = \frac{d}{du}\big(F_U(u)\big) = 6\lambda e^{-2\lambda u} - 12\lambda e^{-4\lambda u} + 6\lambda e^{-6\lambda u} = 6\lambda e^{-2\lambda u}\big(1 - 2e^{-2} + e^{-3}\big)
$$

3.68) Suppose that a queue has n servers and that the length of time to complete a job is an exponential random variable. If a job is at the top of the queue and will be handled by the next available server, what is the distribution of the waiting time until service? What is the distribution of the waiting time until service of the next job in the queue?

Call the independent and identically distributed random variables for the time needed to complete the job $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$. The waiting time until service is given by the minimum of these values, call it V. If V is the minimum, then $V \geq v$ only when $X_1, \ldots, X_n \geq v$. Using the same process as the question above, the distribution for V is given by $f_v(v) = nf_X(v)\big(1 - F_X(v)\big)^{n-1} = n\big(\lambda e^{-\lambda x}\big)\left[1 - \big(1 - e^{-\lambda x}\big)\right]^{n-1} = n\big(\lambda e^{-\lambda x}\big)\left[e^{-\lambda x}\right]^{n-1} = n\lambda e^{-\lambda x n}$. So the wait time until service follows a $\text{Exp}(\lambda n)$ distribution.

The distribution of the wait time until service of the next job is related to the second shortest time one of the n servers finishes their existing job. So we want the 2^{nd} order statistic ($2 - 1 = 1$ samples are smaller, $n - 1$ samples are larger). A counting argument gives this distribution as:

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \frac{n!}{(k-1)!(n-k)!} f(x)F^{k-1}(x)\left[1 - F(x)\right]^{n-k} \\
f_{X_{(2)}}(x) &= \frac{n!}{(n-2)!} f(x)F(x)\left[1 - F(x)\right]^{n-2} \\
f_{X_{(2)}}(x) &= (n^2 - n)(\lambda e^{-\lambda x})(1 - e^{-\lambda x})\left[1 - (1 - e^{-\lambda x})\right]^{n-2} \\
f_{X_{(2)}}(x) &= (n^2 - n)(\lambda e^{\lambda x(1-n)})(1 - e^{-\lambda x}) \\
f_{X_{(2)}}(x) &= (n^2 - n)(\lambda e^{\lambda x(1-n)} - \lambda e^{-\lambda x n}) \\
f_{X_{(2)}}(x) &= (n^2 - n)\lambda e^{-\lambda x n}(e^{\lambda x} - 1)
\end{aligned}
$$

4.72) An item is present in a list of n items with probability p; if it is present, its position in the list is uniformly distributed. A computer program searches through the list sequentially. Find the expected number of items searched through before the program terminates.

Label the list $1, \ldots, n$. The probability the program terminates on the 1st item from the list is the probability that both the item is in the list (p) and the position in the list is the 1st (since uniform, $\frac{1}{n}$), a total probability of $\frac{p}{n}$. This argument repeats for through the first $n-1$ items in the list. The program terminates after n items if either the item is both in the list and in the last position ($\frac{p}{n}$) or if it is not in the list $(1-p)$, a total probability of $\frac{p}{n} + (1-p)$.

So the expected value is:

$$\sum_{i=1}^{n-1} \frac{pi}{n} + \left[n \cdot \left((\frac{p}{n} + (1-p)\right)\right] = \sum_{i=1}^{n} \frac{pi}{n} + [n(1-p)]$$
$$= \frac{p}{n}\left(\frac{n^2+n}{2}\right) + n - np = \frac{pn+p}{2} + \frac{2n-2np}{2} =$$
$$\frac{2n+p-np}{2}$$

4.89) Let $X_1, X_2, \ldots, X_n$ be independent normal random variables with means μ_i and variances σ_i^2. Show that $Y = \sum_{i=1}^{n} \alpha_i X_i$, where the α_i are scalars, is normally distributed, and find its mean and variance using moment generating functions.

Recall the moment generating function of a normal distribution X with mean μ and variance σ^2 is $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Then we calculate Y as follows:

$$M_Y(t) = \mathbb{E}(e^{tY}) = \mathbb{E}\left(e^{t\sum_{i=1}^{n}\alpha_i X_i}\right) = \mathbb{E}\big(e^{t\alpha_1 X_1}\big)\mathbb{E}\big(e^{t\alpha_2 X_2}\big)\cdots\mathbb{E}\big(e^{t\alpha_n X_n}\big) \quad \text{By independence}$$
$$= M_{X_1}(t\alpha_1)M_{X_2}(t\alpha_2)\cdots M_{X_n}(t\alpha_n)$$
$$= \left[e^{\mu_1(t\alpha_1)+\frac{1}{2}\sigma_1^2(t\alpha_1)^2}\right]\left[e^{\mu_2(t\alpha_2)+\frac{1}{2}\sigma_1^2(t\alpha_2)^2}\right]\cdots\left[e^{\mu_n(t\alpha_n)+\frac{1}{2}\sigma_1^2(t\alpha_n)^2}\right]$$
$$= e^{\left[\left(\sum_{i=1}^{n}\mu_i\alpha_i\right)t + \frac{1}{2}t^2\left(\sum_{i=1}^{n}\sigma_i^2\alpha_i^2\right)\right]} \implies Y \sim N\left(\sum_{i=1}^{n}\mu_i\alpha_i, \sum_{i=1}^{n}\sigma_i^2\alpha_i^2\right)$$

1) Let $\boldsymbol{Y_i} \overset{\text{iid}}{\sim}$ Gamma$(\boldsymbol{\alpha, \lambda})$. Find the exact distribution of $\hat{\boldsymbol{Y}}$, then use the delta method to approximate the mean and variance of $\boldsymbol{g(\hat{Y}) = \ln(\hat{Y})}$.

The moment generating function for a gamma distribution x with parameters α and λ is $M_X(t) = \left(\frac{1}{1-\frac{t}{\lambda}}\right)^{\alpha}$ for $t < \alpha$. Then we calculate the exact distribution of $\hat{Y}$ following the same process as above:

$$\begin{aligned} M_{\hat{Y}} &= \mathbb{E}(e^{t\hat{Y}}) = \mathbb{E}\left(e^{\frac{t}{n}\sum_{i=1}^{n} X_i}\right) = \mathbb{E}(e^{\frac{t}{n}Y_1})\mathbb{E}(e^{\frac{t}{n}Y_2})\cdots\mathbb{E}(e^{\frac{t}{n}Y_n}) && \text{By independence} \\ &= M_{Y_1}\left(\frac{t}{n}\right) M_{Y_2}\left(\frac{t}{n}\right)\cdots M_{Y_n}\left(\frac{t}{n}\right) = \left[M_Y\left(\frac{t}{n}\right)\right]^n && \text{Since distributions are identical} \\ &= \left[\left(\frac{1}{1-\frac{t/n}{\lambda}}\right)^{\alpha}\right]^n = \left(\frac{1}{1-\frac{t}{n\lambda}}\right)^{n\alpha} \implies \hat{Y} \sim \text{Gamma}(n\alpha, n\lambda) \end{aligned}$$

The delta method is used to find approximate values of interest for transformed distributions. Since for linear transformations $Y = aX + b$ we have $\mathbb{E}(Y) = a\mathbb{E}(X) + b$ and $\mathbb{V}(Y) = a^2\mathbb{V}(X)$, we can use the first order Taylor Series approximation about a generic transformation $g(X)$ to give a linear approximation for the mean and variance of the transformation. See $g(X) = \sum_{n=0}^{\infty} \frac{g^{(n)(\mu_x)}}{n!}\left(X - \mu_X\right)^n$ so $g(X) \approx g'(\mu_X)X + (g(\mu_X) - g'(\mu_X)\mu_X)$ and then $\mathbb{E}(g(X)) \approx g(\mu_X)$ and $\mathbb{V}(g(X)) \approx [g'(\mu_X)]^2\,\mathbb{V}(X)$.

The expected value of a generic gamma distribution is α/λ and the variance is α/λ^2. From the above, we know the parameters for $\hat{Y}$ are $n\alpha$ and $n\lambda$ respectively. So the approximate mean and variance of $\ln(\hat{Y})$ is calculated as $\mathbb{E}(\ln(X)) \approx \ln(n\alpha/n\lambda) = \ln(\alpha/\lambda)$ and $\mathbb{V}(\ln(X)) \approx \left[g'\left(\frac{n\alpha}{n\lambda}\right)\right]^2 \frac{n\alpha}{(n\lambda)^2} = \left[\frac{\lambda}{\alpha}\right]^2 \frac{\alpha}{n\lambda^2} = \frac{1}{n\alpha}$.

6 Limit Theory

6.1 Definitions

Definition 6.1. Random Sample (RS): $Y_1, Y_2, \ldots, Y_n$ are a random sample if they are independent and identically distributed.

Definition 6.2. Statistic: Functions of a random sample that don't involve unknown parameters. Popular statistics include the **sample mean**, $\hat{Y_n} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, **unbiased sample variance**, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2$

Definition 6.3. Sample Distribution: The distribution of a statistic.

Definition 6.4. Asymptotic Inference: In contrast to exact inference, which is true for any size n, asymptotic inference uses large-samples to approximate distributions.

Definition 6.5. Convergence: Since random variables aren't deterministic, we can't use normal calculus definitions for convergence. Instead, we use three major types of convergence. The strongest type of convergence is "**almost surely**" (that is, with probability 1). This type of convergence is sufficient to show **convergence in probability**. We say a sequence of random variables Y_i converges in probability to a random variable Y, and denote it $Y_i \xrightarrow{\text{P}} Y$, provided for any $\varepsilon > 0$, $\lim_{n\to\infty} P\left(|Y_n - Y| < \varepsilon\right) = 1$ (or equivalently $\lim_{n\to\infty} P\left(|Yn - Y| \geq \varepsilon\right) = 0$). Convergence in probability is a stronger result than **convergence in distribution**. We say a sequence of random variables Y_i with CDFs F_{Y_i} converges in distribution to a random variable Y, and denote it $Y_i \xrightarrow{\text{D}} Y$, if $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ (or equivalently $\lim_{n\to\infty} M_{Y_n}(t) = M_Y(t)$).

6.2 Theorems And Examples

Theorem 6.1. Weak Law Of Large Numbers (WLLN): For independent random variables with known mean, the sample mean converges probabilitistically to the mean as the number of selections grows large. More formally, for a sequence of independent Y_i with $\mathbb{E}(Y_i) = \mu$, and $\mathbb{V}(Y_i) = \sigma^2$, $\hat{Y_n} = \frac{1}{n}\sum\limits_{i=1}^{n} Y_i \xrightarrow{\text{P}} \mu$

Proof. First compute the mean and variance of $\bar{Y_n}$ using independence when needed.

$$\mathbb{E}\left(\bar{Y_n}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mathbb{E}(Y_i)\right) = \frac{1}{n}\left(\sum_{i=1}^{n}\mu\right) = \frac{1}{n}\cdot n\mu = \mu$$
$$\mathbb{V}\left(\bar{Y_n}\right) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{V}(Y_i)\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\sigma^2\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Next, apply Theorem 2.4 (Chebyshev's Inequality, Page 7). Let $\varepsilon > 0$ be given. As $n \to \infty$, we see:

$$P\left(\left|\hat{Y_n} - \mu\right| > \varepsilon\right) = \frac{\mathbb{V}(\hat{Y_n})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

as desired. ■

Theorem 6.2. Continuity Theorems: For continuous g, if Y_i converges to Y, then $g(Y_i)$ converges to $g(Y)$.

Theorem 6.3. Monte Carlo Integration: For definite integrals that necessitate numerical methods, say $\int_a^b g(x)\,dx = I(g)$, then a technique to solve the integral is to generate n random samples from a standard uniform, evaluate g at each observation, and find the average.

Proof. Let $X_i \overset{\text{iid}}{\sim} U(a,b)$. Then $f_X(x) = \frac{1}{b-a}$ for $x \in \{0,1\}$ and zero otherwise. We have $\mathbb{E}(g(x)) = \int_a^b g(x)\frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^b g(x)\,dx$ and apply the WLLN to recall $\frac{1}{n}\sum\limits_{i=1}^{n} g(x_i) = \mathbb{E}(g(x))$. Putting the two together, we observe $\frac{1}{b-a}\sum\limits_{i=1}^{n} g(x) \xrightarrow{\text{p}} \int_a^b g(x)\,dx$ ■

Theorem 6.4. Central Limit Theorem (CLT): For iid random samples, an approximating distribution for the sample mean is a standard normal. More formally, for iid Y_i with finite variance σ^2, $\frac{\hat{Y_n}-\mu}{\sigma/\sqrt{n}} \xrightarrow{\text{d}} N(0,1)$ (or $\hat{Y_n} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$, or $\sqrt{n}\left(\hat{Y_n} - \mu\right) \xrightarrow{\text{d}} N(0,\sigma^2)$).

Proof. We want to show that for iid Y_i with mean μ and variance σ^2, that $\frac{\hat{Y_n}-\mu}{\sigma\sqrt{n}}$ converges in distribution to a standard normal. Rewrite $\frac{\hat{Y}-\mu}{\sigma\sqrt{n}}$ as $\sqrt{n}\frac{\hat{Y_n}-\mu}{\sigma}$. To make it easier for later, we further write:

$$\sqrt{n}\frac{\hat{Y}_n - \mu}{\sigma} = \sqrt{n}\left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n} Y_i - \frac{1}{n}\sum\limits_{i=1}^{n}\mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Y_i - \mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i$$

We compute the MGF as:

$$\begin{aligned}
M_{\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} Z_i}(t) &= \mathbb{E}\left(e^{\frac{t}{\sqrt{n}}\sum\limits_{i=1}^{n} Z_i}\right) = \mathbb{E}\left(e^{\frac{t}{\sqrt{n}}Z_1}e^{\frac{t}{\sqrt{n}}Z_2}\cdots e^{\frac{t}{\sqrt{n}}Z_n}\right) \\
&= \mathbb{E}\left(e^{\frac{t}{\sqrt{n}}Z_1}\right)\mathbb{E}\left(e^{\frac{t}{\sqrt{n}}Z_2}\right)\cdots\mathbb{E}\left(e^{\frac{t}{\sqrt{n}}Z_n}\right) && \text{By independence} \\
&= \left[M_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n && \text{By identical distributions}
\end{aligned}$$

Recall that from the definition of MGF, we have:

$$\begin{aligned}
&M_z^{(k)}(0) = \mathbb{E}\left(Z^k\right) && \text{Definition of MGF} \\
&\mathbb{E}\left(Z^0\right) = 1, \mathbb{E}\left(Z^1\right) = 0, \mathbb{E}\left(Z^2\right) = 1 && \text{From how } Z \text{ is defined}
\end{aligned}$$

We next use Taylor's Series to expand the MGF about zero:

$$\begin{aligned}
M_z(t/\sqrt{n}) &= \sum_{n=0}^{\infty}\frac{M_Z^{(n)}(t/\sqrt{n})}{n!}(t/\sqrt{n} - 0)^n && \text{Definition of Taylor Expansion} \\
&= 1 + 0 + \frac{t^2/n}{2!} + \cdots && \text{Using above expectations} \\
&= 1 + 0 + \frac{t^2/n}{2!} + R_Y(t/\sqrt{n}) && \text{Label the remainder}
\end{aligned}$$

Note that the remainder goes to zero as n goes to infinity. Then Plugging this in from the above, we have:

$$\begin{aligned}
\lim_{n\to\infty} M_z(t/\sqrt{n}) &= \lim_{n\to\infty}\left[M_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n = \lim_{n\to\infty}\left[1 + \frac{n\left(\frac{t^2/n}{2!} + R_Y(t/\sqrt{n})\right)}{n}\right]^n \\
&= \lim_{n\to\infty}\left[1 + \frac{\left(\frac{t^2}{2} + R_Y(t/\sqrt{n})\right)}{n}\right]^n = \lim_{n\to\infty}\left[1 + \frac{\frac{t^2}{2}}{n}\right]^n \\
&= e^{t^2/2}
\end{aligned}$$

The MGF of a $N(\mu, \sigma^2)$ is $e^{\mu t + (\sigma^2)\frac{t^2}{2}}$, so this proves $\hat{Y}_n$ is approximated by a $N(0, 1)$. ∎