Solutions to Selected Problems In: Optimal Statistical Decisions by William DeGroot

John L. Weatherwax*

May 22, 1997

*wax@alum.mit.edu

Chapter 2 (Experiments, Sample Spaces, and Probability)

Problem Solutions

Problem 5 (lemmas of probability distributions)

Part (d): This results is known as Boole's inequality. We begin by decomposing the countable union of events A_i

$$A_1 \cup A_2 \cup A_3 \ldots$$

into a countable union of disjoint events C_j. Define these disjoint events as

$$\begin{aligned}
C_1 &= A_1 \\
C_2 &= A_2 \backslash A_1 \\
C_3 &= A_3 \backslash (A_1 \cup A_2) \\
C_4 &= A_4 \backslash (A_1 \cup A_2 \cup A_3) \\
&\vdots \\
C_j &= A_j \backslash (A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_{j-1})
\end{aligned}$$

Then by construction

$$A_1 \cup A_2 \cup A_3 \cdots = C_1 \cup C_2 \cup C_3 \cdots ,$$

and the C_j's are disjoint events, so that we have (by part (a) of this problem)

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \cdots) = \Pr(C_1 \cup C_2 \cup C_3 \cup \cdots) = \sum_j \Pr(C_j) \, .$$

Since $\Pr(C_j) \leq \Pr(A_j)$ (by part (c) of this problem), for each j, this sum is bounded above by

$$\sum_j \Pr(A_j) \, ,$$

and Boole's inequality is proven.

Problem 6 (the probability that at least one will fail)

Let p denote the probability of failing so that from the given problem we have that $p = 0.01$. Then the probability that at least one component fails is

$$\sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1 - (1-p)^n = 1 - 0.99^n \, ,$$

or the complement of the probability that all components are functional.

Problem 7 (a secretary with letters)

Let E_1, E_2, E_3, E_4, and E_5 be the events that letter $1, 2, 3, 4$ and 5 are placed in their *correct* envelope. Then we are asked about the probability that no letter is placed in its correct envelope or complement of the probability of the event

$$E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 .$$

Thus our probability is given by

$$1 - P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5) .$$

To evaluate $P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5)$ we will use the inclusion/exclusion identity which in this case is given by

$$\begin{aligned} P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5) &= \sum_{i=1}^{5} P(E_i) - \sum_{i<j} P(E_i E_j) + \sum_{i<j<k} P(E_i E_j E_k) \\ &- \sum_{i<j<k<l} P(E_i E_j E_k E_l) + P(E_1 E_2 E_3 E_4 E_5) . \end{aligned}$$

Now each of these joint events is easy to evaluate since they do not depend on the specific values for their indices. Specifically, each the joint events can be evaluated by conditioning on earlier events. For example, for two events we can condition as follows $P(E_i E_j) = P(E_i|E_j)P(E_j)$. Using rules like this we compute

$$\begin{aligned} P(E_i) &= \frac{1}{5} \\ P(E_i E_j) &= \frac{1}{5} \cdot \frac{1}{4} = \frac{1}{20} \\ P(E_i E_j E_k) &= \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{60} \\ P(E_i E_j E_k E_l) &= \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{120} \\ P(E_1 E_2 E_3 E_4 E_5) &= \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{120} . \end{aligned}$$

In addition, since each term in the summations above is a constant (independent of its indices) we can compute the common value and multiply by the number of terms in each sum. Specifically the terms with p events have $\binom{5}{p}$ terms in their summations. With these results we can evaluate the probability of the above union. We find that

$$\begin{aligned} P(\cup_{i=1}^{5} E_i) &= \binom{5}{1}\frac{1}{5} - \binom{5}{2}\frac{1}{20} + \binom{5}{3}\frac{1}{60} - \binom{5}{4}\frac{1}{120} + \binom{5}{5}\frac{1}{120} \\ &= 1 - \frac{1}{2} + \frac{1}{6} - \frac{1}{24} + \frac{1}{120} . \end{aligned}$$

so that our desire probability of no matches is given by

$$1 - P(\cup_{i=1}^{5} E_i) = \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} = \frac{11}{30} \approx 0.3667 .$$

Problem 9 (chaining intersections)

This result follows for the two set case $P\{A \cap B\} = P\{A|B\}P\{B\}$ by grouping the sequence of A_i's in the appropriate manner. For example by grouping the intersection as

$$A_1 \cap A_2 \cap \cdots \cap A_{k-1} \cap A_k = (A_1 \cap A_2 \cap \cdots \cap A_{k-1}) \cap A_k$$

we can apply the two set result to obtain

$$P\{A_1 \cap A_2 \cap \cdots \cap A_{k-1} \cap A_k\} = P\{A_k | A_1 \cap A_2 \cap \cdots \cap A_{k-1}\} \, P\{A_1 \cap A_2 \cap \cdots \cap A_{k-1}\} \, .$$

Continuing now to peal A_{k-1} from the set $A_1 \cap A_2 \cap \cdots \cap A_{k-1}$ we have the second probability above equal to

$$P\{A_1 \cap A_2 \cap \cdots \cap A_{k-2} \cap A_{k-1}\} = P\{A_{k-1} | A_1 \cap A_2 \cap \cdots \cap A_{k-2}\} P\{A_1 \cap A_2 \cap \cdots \cap A_{k-2}\} \, .$$

Continuing to peal off terms from the back we eventually obtain the requested expression i.e.

$$\begin{aligned}
P\{A_1 \cap A_2 \cap \cdots \cap A_{k-1} \cap A_k\} &= P\{A_k | A_1 \cap A_2 \cap \cdots \cap A_{k-1}\} \\
&\times \; P\{A_{k-1} | A_1 \cap A_2 \cap \cdots \cap A_{k-2}\} \\
&\times \; P\{A_{k-2} | A_1 \cap A_2 \cap \cdots \cap A_{k-3}\} \\
&\vdots \\
&\times \; P\{A_3 | A_1 \cap A_2\} \\
&\times \; P\{A_2 | A_1\} \\
&\times \; P\{A_1\} \, .
\end{aligned}$$

If some subset of the intersection has zero probability. Then to show that the entire intersection will have zero probability condition on the intersection that has zero probability. For example, assuming that $P\{\cap_{i=1}^{j} A_i\} = 0$, then grouping the entire intersection as follows

$$\cap_{i=1}^{k} A_i = \left(\cap_{i=1}^{j} A_i\right) \cap \left(\cap_{i=j+1}^{k} A_i\right)$$

We can condition the probability of the entire intersection on the zero intersection set $\cap_{i=1}^{j} A_i$ as follows

$$P\{\cap_{i=1}^{k} A_i\} = P\{\cap_{i=j+1}^{k} A_i | \cap_{i=1}^{j} A_i\} P\{\cap_{i=j+1}^{k} A_i\} \, .$$

This equals zero because $P\{\cap_{i=j+1}^{k} A_i\} = 0$.

Problem 10 (gambling with a fair coin)

Let F denote the event that the gambler is observing results from a fair coin. Also let O_1, O_2, and O_3 denote the three observations made during our experiment. We will assume that before any observations are made the probability that we have selected the fair coin is $1/2$.

Part (a): We desire to compute $P(F|O_1)$ or the probability we are looking at a fair coin given the first observation. This can be computed using Bayes' theorem. We have

$$\begin{aligned} P(F|O_1) &= \frac{P(O_1|F)P(F)}{P(O_1|F)P(F) + P(O_1|F^c)P(F^c)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{2}\right)}{\frac{1}{2}\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right)} = \frac{1}{3}. \end{aligned}$$

Part (b): With the second observation and using the "posteriori's become priors" during a recursive update we now have

$$\begin{aligned} P(F|O_2, O_1) &= \frac{P(O_2|F, O_1)P(F|O_1)}{P(O_2|F, O_1)P(F|O_1) + P(O_2|F^c, O_1)P(F^c|O_1)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{3}\right)}{\frac{1}{2}\left(\frac{1}{3}\right) + 1\left(\frac{2}{3}\right)} = \frac{1}{5}. \end{aligned}$$

Part (c): In this case because the two-headed coin cannot land tails we can immediately conclude that we have selected the fair coin. This result can also be obtained using Bayes' theorem as we have in the other two parts of this problem. Specifically we have

$$\begin{aligned} P(F|O_3, O_2, O_1) &= \frac{P(O_3|F, O_2, O_1)P(F|O_2, O_1)}{P(O_3|F, O_2, O_1)P(F|O_2, O_1) + P(O_3|F^c, O_2, O_1)P(F^c|O_2, O_1)} \\ &= \frac{\frac{1}{2}\left(\frac{1}{5}\right)}{\frac{1}{2}\left(\frac{1}{5}\right) + 0} = 1. \end{aligned}$$

Verifying what we know must be true.

Problem 11 (three different machines)

Let A, B, and C be the events that our items are produced by machine A, B, and C respectively. Then we are told that $P(A) = 0.2$, $P(B) = 0.3$, and $P(C) = 0.5$. Let D be the event that the selected item is defective. Then in the problem formulation we are told that $P(D|A) = 0.04$, $P(D|B) = 0.03$, and $P(D|C) = 0.01$.

Part (a): We are asked to compute $P(A|D)$, $P(B|D)$, and $P(C|D)$. Using Bayes' rule we find that

$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \\ &= (0.04)(0.2) + (0.03)(0.3) + (0.01)(0.5) = 0.022. \end{aligned}$$

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{(0.04)(0.2)}{0.022} = 0.363.$$

In the same way we find that $P(B|D) = 0.409$ and $P(C|D) = 0.227$. Thus the machine from which this defective item most likely came from is B.

Problem 12 (three coins in a box)

Let C_1, C_2, C_3 be the event that the first, second, and third coin is chosen and flipped. Here "first" means the coin is two-headed, "second" means that the coin in two-tailed, and "third" means that the coin is fair. Then let H be the event that the flipped coin showed heads. Then we would like to evaluate $P(C_1|H)$. Using Bayes' rule we have

$$P(C_1|H) = \frac{P(H|C_1)P(C_1)}{P(H)}.$$

We compute $P(H)$ first. We find conditioning on the the coin selected that

$$\begin{aligned} P(H) &= \sum_{i=1}^{3} P(H|C_i)P(C_i) = \frac{1}{3}\sum_{i=1}^{3} P(H|C_i) \\ &= \frac{1}{3}\left(1+0+\frac{1}{2}\right) = \frac{1}{2}. \end{aligned}$$

Then $P(C_1|H)$ is given by

$$P(C_1|H) = \frac{1(1/3)}{(1/2)} = \frac{2}{3}.$$

Problem 15 (sums of binomial coefficients)

For both of these problems we recall the binomial theorem which is

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Then for the first result if we let $x = 1$ and $y = 1$ then $x+y = 2$ and the sum above becomes

$$2^n = \sum_{k=0}^{n} \binom{n}{k},$$

which is the first identity. To prove the second identity if we let $x = -1$ and $y = 1$ then $x + y = 0$ and the sum above then becomes

$$0 = \sum_{k=0}^{n} \binom{n}{k} (-1)^k,$$

which is the second identity.

Problem 16 (choosing r from $x+1$ by drawing subsets of size $r-1$)

Part (a): To show that

$$\binom{x+1}{r} = \binom{x}{r} + \binom{x}{r-1},$$

we consider the left hand side of this expression which represents the number of ways we can choose a subset of size r from $x+1$ objects. Consider this group of $x+1$ objects with one object specified as distinguished or "special". Then the number of ways to select r objects from $x+1$ can be decomposed into two mutually distinct occurrences. The times when this "special" object *is* selected in the subset of size r and the times when its *not*. When it is *not* selected in the subset of size r we are specifying our r subset elements from the x remaining elements giving $\binom{x}{r}$ total subsets in this case. When it *is* selected into the subset of size r we have to select $r-1$ other elements from the x remaining elements, giving $\binom{x}{r-1}$ additional subsets in this case. Summing the counts from these two occurrences we have that $\binom{x+1}{r}$ can be written as the following

$$\binom{x+1}{r} = \binom{x}{r} + \binom{x}{r-1}.$$

We now present an analytic proof of the above. Considering the right hand side of our original expression, we have

$$\begin{aligned}
\binom{x}{r} + \binom{x}{r-1} &= \frac{x!}{(x-r)!r!} + \frac{x!}{(x-r+1)!(r-1)!} \\
&= \frac{(x+1)!}{(x-r)!r!(x+1)} + \frac{(x+1)!}{(x-r+1)!(r-1)!(x+1)} \\
&= \frac{(x+1)!}{(x+1-r)!r!}\left(\frac{x+1-r}{x+1}\right) + \frac{(x+1)!}{(x-r+1)!r!}\left(\frac{r}{x+1}\right) \\
&= \frac{(x+1)!}{(x-r+1)!r!}\left(\frac{x+1-r}{x+1} + \frac{r}{x+1}\right) \\
&= \binom{x+1}{r},
\end{aligned}$$

and the result is proven.

Part (b): To show the given sum consider the identity given above which can be written (with the variable r replaced with k)

$$\binom{x}{k-1} = \binom{x+1}{k} - \binom{x}{k}.$$

Summing this expression for $x = k-1$ to $x = r+k-1$ we see that

$$\begin{aligned}
\sum_{x=k-1}^{r+k-1} \binom{x}{k-1} &= \left(\binom{k}{k} - \binom{k-1}{k}\right) + \left(\binom{k+1}{k} - \binom{k}{k}\right) + \cdots \\
&+ \left(\binom{r+k-1}{k} - \binom{r+k-2}{k}\right) + \left(\binom{r+k}{k} - \binom{r+k-1}{k}\right) \\
&= -\binom{k-1}{k} + \binom{r+k}{k} \\
&= \binom{r+k}{k}
\end{aligned}$$

where we have used the convention that $\binom{k-1}{k} = 0$.

Chapter 3 (Random variables and distribution functions)

Problem Solutions

Problem 11 (the probability density function for $Y = aX + b$)

We begin by computing the cumulative distribution function of the random variable Y as

$$\begin{aligned} F_Y(y) &= P\{Y \le y\} \\ &= P\{aX + b \le y\} \\ &= P\{X \le \frac{y-b}{a}\} \\ &= F_X(\frac{y-b}{a}) . \end{aligned}$$

Taking the derivative to obtain the distribution function for Y we find that

$$f_Y(y) = \frac{dF_Y}{dy} = F_X'(\frac{y-b}{a})\frac{1}{a} = \frac{1}{a} f_X(\frac{y-b}{a}) .$$

Problem 12 (the probability density function for $Y = X^2$)

Assume that X is distributed with a density function $f_X(x)$. Define the random variable $Y = X^2$, then this problem asks to find the distribution function for Y. We can calculate this by first calculating the cumulative distribution function for Y, i.e.

$$\begin{aligned} F_Y(a) &= Pr\{Y \le a\} \\ &= Pr\{X^2 \le a\} \\ &= Pr\{-\sqrt{a} \le X \le +\sqrt{a}\} \\ &= \int_{-\sqrt{a}}^{\sqrt{a}} f_X(\xi) d\xi . \end{aligned}$$

Thus the distribution function for Y is given by the derivative of the above expression. This derivative can be computed as follows

$$\begin{aligned} f_Y(a) &= \frac{dF_Y}{da} = f_X(\sqrt{a})\frac{d(\sqrt{a})}{da} - f_X(-\sqrt{a})\frac{d(-\sqrt{a})}{da} \\ &= f_X(\sqrt{a})\frac{1}{2\sqrt{a}} + f_X(-\sqrt{a})\frac{1}{2\sqrt{a}} \\ &= \frac{f_X(\sqrt{a}) + f_X(-\sqrt{a})}{2\sqrt{a}} . \end{aligned}$$

Problem 13 (some univariate distribution function from the joint)

We are given that X_1 and X_2 are jointly distributed with a distribution function $f_{X_1,X_2}(x_1, x_2)$ and we first desire to compute the univariate distribution for the sum of X_1 and X_2. To find this lets consider the cumulative distribution function for the random variable $X_1 + X_2$, specifically we have (defining $Z = X_1 + X_2$)

$$\begin{aligned} F_Z(a) &= Pr\{Z \le a\} \\ &= Pr\{X_1 + X_2 \le a\} \\ &= \int_{x_1=-\infty}^{+\infty} \int_{x_2=-\infty}^{a-x_1} f_{X_1,X_2}(x_1, x_2)\, dx_2\, dx_1 \,. \end{aligned}$$

This integral can be derived by assuming that a is positive and drawing the line $x_1 + x_2 = a$, in the (X_1, X_2) plane. The desired probability is the "area" beneath this line. Once this integral is evaluated (its evaluation depends on the specific functional form of $f_{X_1,X_2}(x_1, x_2)$) the density function for $f_Z(a)$ is given by taking the derivative of the above cumulative density function

$$f_Z(a) = \frac{dF_z(a)}{da} \,.$$

Problem 17 (the probability integral transformation)

Part (a): If $Y = F(X)$ then the distribution function of Y is given by

$$\begin{aligned} F_Y(a) &= P\{Y \le a\} \\ &= P\{F(X) \le a\} \\ &= P\{X \le F^{-1}(a)\} \\ &= F(F^{-1}(a)) = a \,. \end{aligned}$$

Thus $f_Y(a) = \frac{dF_Y}{da} = 1$, showing that Y is a uniform random variable.

Chapter 4 (Some special univariate distributions)

Notes on sections in the text

Notes on the hypergeometric distribution

When X is given by a hypergeometric distribution it has a p.d.f. given by

$$f(x|A, B, n) = \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}, \tag{1}$$

which represents the probability of getting x type A objects from a total of $A + B$ objects, when we draw n objects. Note that the *range* of the values possible for the random variable X is determined by the values of A, B, and n. Since we only have A total objects of type A the largest X value can be is $\min(A, n)$. At the same time we have B objects of type B and if we happen to draw *all* of them we will have used up B of our total n draws on B objects. Thus the remaining $n - B$ object that we draw must all be objects of type A. Thus the minimum value of our random variable X must be larger than $\max(0, n - B)$. In total then we have bounds on X when X is a hypergeometric random variable with parameters A, B, and n of the form

$$\min(0, n - B) \le X \le \max(A, n) . \tag{2}$$

Notes on the Gamma distribution

The book makes the claim (but provides no proof) that when $X_1, \ldots, X_n$ are random samples from a normal distribution with mean μ and variance σ^2, and when

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 ,$$

then S^2 has a χ^2 distribution with $n - 1$ degrees of freedom. A relatively simple proof of this fact can be found in Appendix A (chi-squared distribution) of [4].

Notes on the t-distribution

If X_i are normal random variables with mean μ and variance σ^2 then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has a mean μ and a variance $\frac{\sigma^2}{n}$. Thus the variable $\frac{n^{1/2}(\bar{X}-\mu)}{\sigma}$ is standard normal. We are also told in the section on the gamma distribution that the normalized sample standard deviation or

$$S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \bar{x})^2 ,$$

is χ^2 with $n-1$ degrees of freedom. Then to compute a t distributed random variable using the "normal over a χ^2 rule we can construct

$$\frac{\frac{n^{1/2}(\bar{X}-\mu)}{\sigma}}{(S^2/(n-1))^{1/2}} = \frac{n^{1/2}(\bar{X}-\mu)}{\sigma\left[\frac{1}{\sigma^2(n-1)}\sum_{i=1}^n (X_i-\bar{X})^2\right]^{1/2}} = \frac{n^{1/2}(\bar{X}-\mu)}{\left[\frac{1}{(n-1)}\sum_{i=1}^n (X_i-\bar{X})^2\right]^{1/2}},$$

is a t distributed random variable with $n-1$ degrees of freedom as stated in the books equation 3.

If we take $X = \tau^{1/2}(Y-\mu)$ where X is a t distributed random variable with a p.d.f. given by

$$g_X(x) = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)}\left(1+\frac{x^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)}.$$

Then from X we can compute Y to get $Y = \frac{X}{\tau^{1/2}} + \mu$ and the p.d.f. of Y would be given by

$$g_Y(y) = g_X(x(y))\left|\frac{dx}{dy}\right| = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)}\left(1+\frac{\tau(y-\mu)^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)}\tau^{1/2},$$

which is the p.d.f. given by the books equation 5.

Notes on the F-distribution

Since $S_X^2 = \frac{1}{\sigma^2}\sum_{i=1}^m (X_i-\bar{X})^2$ is χ^2 with $m-1$ degrees of freedom and $S_Y^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (Y_i-\bar{Y})^2$ is χ^2 with $m-1$ degrees of freedom the *ratio*

$$\frac{S_X^2/(m-1)}{S_Y^2/(n-1)} = \frac{\frac{1}{m-1}\sum_{i=1}^m (X_i-\bar{X})^2}{\frac{1}{n-1}\sum_{i=1}^n (Y_i-\bar{Y})^2},$$

is the ratio of two χ^2 random variables divided by their degrees of freedom and therefore is given by a F-distribution with parameters $m-1$ and $n-1$.

Exercise Solutions

Exercise 1 (properties of a Bernoulli random variable)

We find

$$E(X) = 1p + 0(1-p) = p,$$

and

$$E(X^2) = 1^2p + 0^2(1-p) = p.$$

Thus

$$\text{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) = pq.$$

The characteristic function $\zeta(t)$ is given by

$$\zeta(t) = E(e^{itX}) = pe^{it1} + qe^{it0} = pe^{it} + q.$$

Exercise 2 (some properties of the binomial random variable)

For a binomial random variable we find its expectation given by $E(X)$

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} = \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} \\
&= n \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=1}^{n} \left(\begin{array}{c} n-1 \\ x-1 \end{array} \right) p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=0}^{n-1} \left(\begin{array}{c} n-1 \\ x \end{array} \right) p^{x} q^{(n-1)-x} \\
&= np \cdot 1 = np\,.
\end{aligned}
$$

Next we need to evaluate $E(X^2)$. We find

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{n} x^2 \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n(n-1)!}{(x-1)!(n-x)!} p^{x-1+1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=1}^{n} (x-1+1) \left(\begin{array}{c} n-1 \\ x-1 \end{array} \right) p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=1}^{n} (x-1) \left(\begin{array}{c} n-1 \\ x-1 \end{array} \right) p^{x-1} q^{(n-1)-(x-1)} + np \sum_{x=1}^{n} \left(\begin{array}{c} n-1 \\ x-1 \end{array} \right) p^{x-1} q^{(n-1)-(x-1)} \\
&= np \sum_{x=2}^{n} (x-1) \left(\begin{array}{c} n-1 \\ x-1 \end{array} \right) p^{x-1} q^{(n-1)-(x-1)} + np \sum_{x=0}^{n-1} \left(\begin{array}{c} n-1 \\ x \end{array} \right) p^{x} q^{(n-1)-x} \\
&= np \sum_{x=2}^{n} \frac{(n-1)(n-2)!}{(x-2)!((n-1)-(x-1))!} p^{x-2+1} q^{(n-2)-(x-2)} + np \\
&= n(n-1)p^2 \sum_{x=2}^{n} \left(\begin{array}{c} n-2 \\ x-2 \end{array} \right) p^{x-2} q^{(n-2)-(x-2)} + np \\
&= n(n-1)p^2 \sum_{x=0}^{n-2} \left(\begin{array}{c} n-2 \\ x \end{array} \right) p^{x} q^{(n-2)-x} + np \\
&= n(n-1)p^2 + np\,.
\end{aligned}
$$

Thus the variance of a binomial random variable is given by combining these two results as

$$
\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2 \\
&= np(1-p) = npq\,. \qquad (3)
\end{aligned}
$$

To show that the characteristic function for a binomial random variable is given by $\zeta(t) = (pe^{it} + q)^n$, it suffices to observe that a binomial random variable can be written as the sum of n independent Bernoulli random variables. Because of this the characteristic function for a binomial random variable is the product of n Bernoulli random variables characteristic functions. From Exercise 1 above, the characteristic function for a Bernoulli random variable is given by $pe^{it} + q$. Thus the characteristic function for our binomial random variable is given by this to the nth power or

$$\zeta(t) = (pe^{it} + q)^n . \tag{4}$$

Another way to derive this result is to compute it directly. We have

$$\begin{aligned} \zeta(t) &= E[e^{itX}] = \sum_{x=0}^{n} \binom{n}{x} e^{itx} p^x q^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^{it}p)^x q^{n-x} \\ &= (pe^{it} + q)^n , \end{aligned}$$

using the binomial expression.

Exercise 3 (summing binomial random variables)

If each X_i is a binomial random variable with parameters p and n_i then it can be represented as the sum of n_i Bernoulli random variables as

$$X_i = V_1 + V_2 + \cdots + V_{n_i} .$$

With this decomposition, the sum $X_1 + X_2 + \cdots X_k$ can then be written as a larger sum

$$V_1 + V_2 + \cdots V_{n_1} + V_{n_1+1} + \cdots + V_{n_1+n_2} + \cdots + V_{n_1+n_2+\cdots+n_{k-1}+1} + \cdots + V_{n_1+n_2+\cdots+n_{k-1}+n_k} ,$$

or the sum of $n_1 + n_2 + \cdots + n_k$ Bernoulli random variables each with the same probability of success p. Thus we recognize this sum as another binomial random variable with parameters $n_1 + n_2 + \cdots + n_k$ and p as claimed.

Alternatively, as the X_i random variables are independent the characteristic function for their *sum* is given by the *product* of the characteristic function for each individual random variable X_i. Thus we have

$$\zeta_{X_1+X_2+\cdots+X_k}(t) = \prod_{i=1}^{k} (pe^{it} + q)^{n_i} = (pe^{it} + q)^{\sum_{i=1}^{k} n_i} ,$$

which we see is the characteristic function for a binomial random variable with parameters $\sum_{i=1}^{k} n_i$ and p as claimed.

Exercise 4 (the limit of the binomial p.d.f is a Poisson p.d.f)

Lets begin by writing the density of a binomial random variable $f(x|n,p)$ as

$$
\begin{aligned}
f(x|n,p) &= \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x} \\
&= \frac{1}{x!}\frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{(n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1}p^x(1-p)^{n-x}\,.
\end{aligned}
$$

To let $n\to\infty$ and $p\to 0$ in such a way that $np\to\lambda$ we will take $p=\frac{\lambda}{n}$. Then $f(x|n,p)$ becomes

$$
\begin{aligned}
f(x|n,p) &= \frac{1}{x!}\frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{(n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x} \\
&= \frac{1}{x!}\frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{(n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1}\frac{\lambda^x}{n^x}\frac{(n-\lambda)^{n-x}}{n^{n-x}} \\
&= \frac{\lambda^x}{x!}\frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{(n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1}\frac{\lambda^x}{n^x}\frac{(n-\lambda)^{n}}{n^{n}}(n-\lambda)^{-x} \\
&= \frac{\lambda^x}{x!}\frac{n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1}{(n-\lambda)(n-\lambda)(n-\lambda)\cdots(n-\lambda)(n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1} \\
&\times \left(1-\frac{\lambda}{n}\right)^{n}\,. \qquad (5)
\end{aligned}
$$

Note that the factorial fraction above can be written (after canceling some terms) as the product

$$
\frac{n}{n-\lambda}\cdot\frac{n-1}{n-\lambda}\cdot\frac{n-2}{n-\lambda}\cdots\frac{n-x+2}{n-\lambda}\cdot\frac{n-x+1}{n-\lambda}\,,
$$

or dividing by n on the "top and bottom" we get

$$
\frac{1}{1-\frac{\lambda}{n}}\cdot\frac{1-\frac{1}{n}}{1-\frac{\lambda}{n}}\cdot\frac{1-\frac{2}{n}}{1-\frac{\lambda}{n}}\cdots\frac{1-\frac{x-2}{n}}{1-\frac{\lambda}{n}}\cdot\frac{1-\frac{x-1}{n}}{1-\frac{\lambda}{n}}\,.
$$

Each of the factors in this product goes to $\frac{1}{1}\to 1$ as $n\to\infty$. Also recall that right-most factor in Equation 5 has the following limit

$$
\left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}\,,
$$

as $n\to\infty$. Using both of these results we have that

$$
f(x|n,p)=\frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}\to\frac{1}{x!}\lambda^x e^{-\lambda}=g(x|\lambda)\,, \qquad (6)
$$

with $\lambda=np$ as we were to show.

Exercise 5 (properties of a Poisson random variable)

If X is a Poisson random variable then from the definition of expectation we have that

$$E[X^n] = \sum_{i=0}^{\infty} i^n e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{i^n\lambda^n}{i!}e^{-\lambda} = \sum_{i=1}^{\infty}\frac{i^n\lambda^i}{i!},$$

since (assuming $n \neq 0$) when $i = 0$ the first term vanishes. Continuing our calculation we can cancel a factor of i and find that

$$\begin{aligned} E[X^n] &= e^{-\lambda}\sum_{i=1}^{\infty}\frac{i^{n-1}\lambda^i}{(i-1)!} = e^{-\lambda}\sum_{i=0}^{\infty}\frac{(i+1)^{n-1}\lambda^{i+1}}{i!} \\ &= \lambda\sum_{i=0}^{\infty}\frac{(i+1)^{n-1}e^{-\lambda}\lambda^i}{i!}. \end{aligned}$$

Now this sum can be recognized as the expectation of the variable $(X+1)^{n-1}$ so we see that

$$E[X^n] = \lambda E[(X+1)^{n-1}]. \tag{7}$$

From the result we have

$$E[X] = \lambda E[1] = \lambda \quad \text{and} \quad E[X^2] = \lambda E[X+1] = \lambda(\lambda+1).$$

Thus the variance of X is given by

$$\text{Var}[X] = E[X^2] - E[X]^2 = \lambda.$$

We find the characteristic function for a Poisson random variable given by

$$\begin{aligned} \zeta(t) &= E[e^{itX}] = \sum_{x=0}^{\infty} e^{itx}\frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^{it}\lambda)^x}{x!} = e^{-\lambda}e^{\lambda e^{it}} \\ &= \exp\{\lambda(e^{it}-1)\}. \end{aligned} \tag{8}$$

Above we explicitly calculated $E(X)$ and $\text{Var}(X)$ but we can also use the above characteristic function to derive them. For example, we find

$$\begin{aligned} E(X) &= \frac{1}{i}\frac{\partial\zeta(t)}{\partial t}\Big|_{t=0} = \frac{1}{i}\exp\{\lambda(e^{it}-1)\}\lambda i e^{it}\Big|_{t=0} \\ &= \lambda e^{it}\exp\{\lambda(e^{it}-1)\}\Big|_{t=0} = \lambda, \end{aligned}$$

for $E(X)$ and

$$\begin{aligned} E(X^2) &= \frac{1}{i^2}\frac{\partial^2\zeta(t)}{\partial t^2}\Big|_{t=0} = \frac{1}{i}\frac{\partial}{\partial t}\left(\lambda e^{it}\exp\{\lambda(e^{it}-1)\}\right)\Big|_{t=0} \\ &= \frac{1}{i}\left[i\lambda e^{it}\exp\{\lambda(e^{it}-1)\} + \lambda e^{it}(\lambda i e^{it})\exp\{\lambda(e^{it}-1)\}\right]\Big|_{t=0} \\ &= \lambda e^{it}\exp\{\lambda(e^{it}-1)\} + \lambda^2 e^{2it}\exp\{\lambda(e^{it}-1)\}\Big|_{t=0} \\ &= \lambda + \lambda^2, \end{aligned}$$

for $E(X^2)$ the same two results as before.

Exercise 6 (the sums of Poisson random variables)

We will prove this result in the case of two Poisson random variables X and Y (with means λ_1 and λ_2) and then just state mathematical induction to derive the requested result in the case of a sum of a finite number of random variables Poisson variables. To begin we note that we can evaluate the distribution of $X+Y$ by computing the characteristic function of $X+Y$. Since X and Y are both Poisson random variables the characteristic functions of $X+Y$ is given by

$$\begin{aligned}\phi_{X+Y}(u) &= \phi_X(u)\phi_Y(u) \\ &= e^{\lambda_1(e^{iu}-1)}e^{\lambda_2(e^{iu}-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^{iu}-1)}.\end{aligned}$$

From the direct connection between characteristic functions to and probability density functions we see that the random variable $X+Y$ is a Poisson random variable with parameter $\lambda_1+\lambda_2$, the sum of the Poisson parameters of the random variables X and Y.

Exercise 7 (the distribution of X given $X+Y$)

Rather than work with X_1 and X_2 we will consider Poisson random variables denoted by X and Y. Then this problem asks for the conditional distribution of X given $X+Y$. Define the random variable Z by $Z = X+Y$. Then from Bayes' rule we find that

$$p(X|Z) = \frac{p(Z|X)p(X)}{p(Z)}.$$

We will evaluate each expression in tern. Now $p(X)$ is the probability density function of a Poisson random variable with parameter λ_1 so $p(X=x) = \frac{e^{-\lambda_1}\lambda_1^x}{x!}$. From problem 6 in this chapter we have that $P(Z=n) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$. Finally to evaluate $p(Z=n|X=x)$ we recognize that this is equivalent to $p(Y=n-x)$, which we can evaluate easily. We have that

$$p(Z=n|X=x) = p(Y=n-x) = \frac{e^{-\lambda_2}\lambda_2^{n-x}}{(n-x)!}.$$

Putting all of these pieces together we find that

$$\begin{aligned}p(X=x|Z=n) &= \left(\frac{e^{-\lambda_2}\lambda_2^{n-x}}{(n-x)!}\right)\left(\frac{e^{-\lambda_1}\lambda_1^x}{x!}\right)\left(\frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}\right) \\ &= \left(\frac{n!}{x!(n-x)!}\right)\frac{\lambda_1^x\lambda_2^{n-x}}{(\lambda_1+\lambda_2)^n} \\ &= \left(\begin{array}{c} n \\ x \end{array}\right)\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-x}.\end{aligned}$$

Defining $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$ and $q = 1-p = \frac{\lambda_2}{\lambda_1+\lambda_2}$ our density above becomes

$$p(X=x|Z=n) = \left(\begin{array}{c} n \\ x \end{array}\right)p^x(1-p)^{n-x},$$

or in words $p(X=x|Z=n)$ is a Binomial random variable with parameters $(n, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

Exercise 8 (an alternative representation for the negative binomial distribution)

Using the definition of $\left(\begin{array}{c} a \\ x \end{array}\right)$ given by

$$\left(\begin{array}{c} a \\ x \end{array}\right) = \frac{\prod_{i=0}^{x-1}(a-i)}{x!}, \tag{9}$$

we see that we can write $\left(\begin{array}{c} r+x-1 \\ x \end{array}\right)$ that appears in the expression for the p.f. of a negative binomial random variable as

$$\begin{aligned}
\left(\begin{array}{c} r+x-1 \\ x \end{array}\right) &= \frac{1}{x!}\prod_{i=0}^{x-1}(r-x-1-i) \\
&= \frac{1}{x!}(r-x-1)(r-x-2)\cdots(r-x-1-(x-2))(r-x-1-(x-1)) \\
&= \frac{1}{x!}(r-x-1)(r-x-2)\cdots(r+1)r\,.
\end{aligned}$$

While we can also consider the expression $\left(\begin{array}{c} -r \\ x \end{array}\right)$ presented in this problem. Here we see that it is equal to

$$\begin{aligned}
\left(\begin{array}{c} -r \\ x \end{array}\right) &= \frac{1}{x!}\prod_{i=0}^{x-1}(-r-i) \\
&= \frac{1}{x!}(-1)^x r(r+1)(r+2)\cdots(r+x-2)(r+x-1)\,.
\end{aligned}$$

Thus the suggested p.f. of $\left(\begin{array}{c} -r \\ x \end{array}\right) p^r(-q)^x$ given by

$$\left(\begin{array}{c} -r \\ x \end{array}\right) p^r(-q)^x = \frac{1}{x!}r(r+1)(r+2)\cdots(r+x-2)(r+x-1) = \left(\begin{array}{c} r+x-1 \\ x \end{array}\right) p^r q^x\,,$$

the same as the expression for the p.f. for a negative binomial random variable proving the equivalence.

Exercise 9 (the characteristic function for the negative binomial distribution)

We can evaluate this using

$$\zeta(t) = E(e^{itX}) = \sum_{x=0}^{\infty}\left(\begin{array}{c} r+x-1 \\ x \end{array}\right) p^r q^x e^{itx}\,,$$

or by using the result from Exercise 8 we can write this as

$$\zeta(t) = E(e^{itX}) = p^r\sum_{x=0}^{\infty}\left(\begin{array}{c} -r \\ x \end{array}\right)(-qe^{it})^x\,.$$

This later sum is the Taylor expansion of the expression $(1-qe^{it})^{-r}$ see [1]. Thus we find

$$\zeta(t) = \frac{p^r}{(1-qe^{it})^r}. \tag{10}$$

To calculate $E(X)$ we have

$$\begin{aligned}
E(X) &= \frac{1}{i}\frac{\partial}{\partial t}\left(\frac{p^r}{(1-qe^{it})^r}\right)\Big|_{t=0} \\
&= \frac{1}{i}\,p^r(-r)(1-qe^{it})^{-r-1}(-qie^{it})\big|_{t=0} \\
&= rp^r qe^{it}(1-qe^{it})^{-r-1}\big|_{t=0} \\
&= \frac{rp^r q}{(1-q)^{r+1}} = \frac{rp^r q}{p^{r+1}} = \frac{rq}{p}.
\end{aligned}$$

For $E(X^2)$ we have

$$\begin{aligned}
E(X^2) &= \frac{1}{i^2}\frac{\partial^2}{\partial t^2}\left(\frac{p^r}{(1-qe^{it})^r}\right)\Big|_{t=0} \\
&= \frac{1}{i}\frac{\partial}{\partial t}\left(rp^r qe^{it}(1-qe^{it})^{-r-1}\right)\big|_{t=0} \\
&= \frac{rp^r q}{i}\left[ie^{it}(1-qe^{it})^{-r-1} + e^{it}(-r-1)(-qie^{it})(1-qe^{it})^{-r-2}\right]\big|_{t=0} \\
&= rp^r q\left[\frac{1}{(1-q)^{r+1}} + \frac{(r+1)q}{(1-q)^{r+2}}\right] \\
&= rp^r q\left[\frac{p}{p^{r+2}} + \frac{(r+1)q}{p^{r+2}}\right] = \frac{rq}{p^2}(p+rq+q) \\
&= \frac{rq(1+rq)}{p^2}.
\end{aligned} \tag{11}$$

Thus the variance for a negative binomial random variable is given by

$$\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - E(X)^2 \\
&= \frac{rq(1+rq)}{p^2} - \frac{r^2q^2}{p^2} = \frac{rq}{p^2}.
\end{aligned} \tag{12}$$

Exercise 10 (summing negative binomial random variables)

We can use the fact that a random variable that is defined as the sum of independent random variables has a characteristic function that is the product of the individual characteristic functions. Since the characteristic function of each X_i is given by

$$\zeta_{X_i}(t) = \left(\frac{p}{1-qe^{it}}\right)^{r_i},$$

The characteristic function of the the sum $\sum_{i=1}^k X_i$ is given by

$$\zeta_{X_1+X_2+\cdots+X_k}(t) = \left(\frac{p}{1-qe^{it}}\right)^{\sum_{i=1}^k r_i},$$

which is the characteristic function of a negative binomial random with parameters $\sum_{i=1}^k r_i$ and p as we were to show.

Exercise 11 (the limiting p.d.f of a negative binomial is Poisson)

If we want $r \to \infty$ and $q \to 0$ in such a way that $rq \to \lambda$, then one way this can be guaranteed to happen is to define q in terms of r as $q = \frac{\lambda}{r}$. Thus in this case we see that as $r \to \infty$ then $q \to 0$. Now consider $f(x|r,p) = \left(\begin{array}{c} r+x-1 \\ x \end{array} \right) p^r q^x$ the probability density for the negative binomial random variable. For q defined as above we have

$$
\begin{aligned}
f(x|r,p) &= \left(\begin{array}{c} r+x-1 \\ x \end{array} \right) p^r \left(\frac{\lambda}{r}\right)^x \\
&= \frac{(r+x-1)(r+x-2)(r+x-3)\cdots(r+2)(r+1)r(r-1)(r-2)\cdots 3\cdot 2\cdot 1}{x!(r-1)(r-2)(r-3)\cdots 3\cdot 2\cdot 1} p^r \frac{\lambda^x}{r^x} \\
&= \frac{\lambda^x}{x!}\left[\frac{(r+x-1)(r+x-2)(r+x-3)\cdots(r+2)(r+1)r}{r\cdot r\cdot r\cdots r\cdot r\cdot r}\right](1-q)^r \\
&= \frac{\lambda^x}{x!}\left[\frac{r+x-1}{r}\cdot\frac{r+x-2}{r}\cdot\frac{r+x-3}{r}\cdots\frac{r+2}{r}\cdot\frac{r+1}{r}\right]\left(1-\frac{\lambda}{r}\right)^r .
\end{aligned}
$$

As $r \to \infty$ each factor in the brackets goes to $+1$ and using the fact that

$$\left(1+\frac{\lambda}{n}\right)^n \to e^{\lambda},$$

we see that our density $f(x|r,p)$ has the limit given limit

$$f(x|r,p) \to \frac{e^{-\lambda}\lambda^x}{x!},$$

as we were to show.

Exercise 12 (X_i has a negative binomial p.d.f)

Warning: I was not able to solve this problem. What follows are some notes on a few things that I tried and where I got stuck. If anyone sees anything wrong with these notes or knows of a different way to work this problem please email me.

Note that if Y is a Poisson random variable with a mean $r\ln(1/p)$ then it has a form like

$$P\{Y=n\} = \frac{e^{-r\ln(1/p)}(r\ln(1/p))^n}{n!} = \frac{p^r(-r\ln(p))^n}{n!}.$$

From the given description of the problem we can compute some values for $P\{Z=z\}$ for some simple values of z, and then verify that the values computed equal the same thing that one gets from the p.d.f of a negative binomial distribution $\binom{-r}{n} p^r(-q)^n$. For example,

$$P\{Z=0\} = P\{Y=0\} = p^r\,,$$

which does equal $\binom{-r}{0} p^r(-q)^0$. Another simple probability to calculate is $P\{Z=1\}$ where we find

$$P\{Z=1\} = P\{Y=1, X_1=1\} = P\{Y=1\}P\{X_1=1\} = p^r(-r\ln(p))\frac{1}{\ln(1/p)}q = rp^r q\,,$$

which again does equal $\binom{-r}{1} p^r(-q)^1$ because

$$\binom{-r}{1} = \frac{\prod_{i=0}^{0}(-r-i)}{0!} = \frac{-r}{1} = -r\,,$$

and so

$$\binom{-r}{1} p^r(-q)^1 = rp^r q\,,$$

We could evaluate $P\{Z=2\}$ in the same way as

$$P\{Z=2\} = P\{Y=1, X_1=2\} + P\{Y=2, X_1=1, X_2=1\}\,,$$

but we stop here.

This discussion motivated attempting to evaluate $P\{Z=k\}$ by conditioning on the value of X_1. We have

$$\begin{aligned}
P\left\{\sum_{i=1}^{Y} X_i = k\right\} &= \sum_{j=1}^{k} P\left\{\sum_{i=1}^{Y} X_i = k|x_1 = j\right\} P\left\{X_1 = j\right\} \\
&= \sum_{j=1}^{k} P\left\{\sum_{i=1}^{Y} X_i = k|x_1 = j\right\} \frac{1}{\ln(1/p)}\frac{q^j}{j} \\
&= \frac{1}{\ln(1/p)}\sum_{j=1}^{k}\frac{q^j}{j} P\left\{\sum_{i=1}^{Y} X_i = k|X_1 = j\right\} \\
&= \frac{1}{\ln(1/p)}\sum_{j=1}^{k}\frac{q^j}{j} P\left\{\sum_{i=2}^{Y} X_i = k-j|X_1 = j\right\} \\
&= \frac{1}{\ln(1/p)}\sum_{j=1}^{k}\frac{q^j}{j} P\left\{\sum_{i=1}^{Y-1} X_i = k-j\right\}\,.
\end{aligned}$$

Thus if we define the function $f(n,k)$ as $f(n,k) = P\left\{\sum_{i=1}^{n} X_i = k\right\}$, we have a recursive relationship for $f(n,k)$ given by

$$f(n,k) = \frac{1}{\ln(1/p)}\sum_{j=1}^{k}\frac{q^j}{j} f(n-1, k-j)\,.$$

This later summation looks like a convolution type sum and maybe there are identities for it. One might be able to solve this equation directly or show by substitution that one solution for $f(n,k)$ is $\left(\begin{array}{c} -k \\ n \end{array}\right) p^k(-q)^n$, which would solve the given problem. When I put in the expression $\left(\begin{array}{c} -k \\ n \end{array}\right) p^k(-q)^n$ into the right-hand-side of the above we obtain

$$\frac{(-q)^{n-1}p^k}{\ln(1/p)} \sum_{j=1}^{k} \frac{q^j p^{-j}}{j} \left(\begin{array}{c} -(k-j) \\ n-1 \end{array}\right).$$

Consider just the sum

$$\sum_{j=1}^{k} \frac{p^j q^{-j}}{j} \left(\begin{array}{c} -(k-j) \\ n-1 \end{array}\right) = \sum_{j=1}^{k} \left(\frac{p}{q}\right)^j \left(\begin{array}{c} -(k-j) \\ n-1 \end{array}\right).$$

Now if we let $r = \frac{p}{q}$ and define a function $F(r)$ as

$$F(r) = \sum_{j=1}^{k} \frac{r^j}{j} \left(\begin{array}{c} -(n-j) \\ n-1 \end{array}\right),$$

the derivative of this expression with respect to r is

$$F'(r) = \sum_{j=1}^{k} r^{j-1} \left(\begin{array}{c} -(n-j) \\ n-1 \end{array}\right) = \sum_{j=1}^{k} r^{j-1} \left(\begin{array}{c} -(n-1-(j-1)) \\ n-1 \end{array}\right) = \sum_{j=0}^{k-1} r^j \left(\begin{array}{c} -(n-1-j) \\ n-1 \end{array}\right)$$

This is were I stopped. I was not able to perform the summation of this sum and show the equivalence. Again, if anyone knows of a way to complete this problem or an alternative method please email me.

Exercise 13 (statistical properties of a hypergeometric random variable)

If X is a hypergeometric random variable then X has a p.d.f. given by Equation 1 and the valid range of values for X given by Equation 2 and zero otherwise. Then using this expression for the p.d.f of X we can compute $E(X)$. We find

$$E(X) = \sum_{X} x f(x|A,B,n) = \frac{1}{\left(\begin{array}{c} A+B \\ n \end{array}\right)} \sum_{X'} x \left(\begin{array}{c} A \\ x \end{array}\right) \left(\begin{array}{c} B \\ n-x \end{array}\right),$$

where the X' notation in the sum means that we don't include the point $x = 0$ in the above sum since it does not affect the expectation any. To simplify this remaining sum recall that

$$\left(\begin{array}{c} A \\ x \end{array}\right) = \frac{A}{x} \left(\begin{array}{c} A-1 \\ x-1 \end{array}\right),$$

so that

$$x \left(\begin{array}{c} A \\ x \end{array}\right) = A \left(\begin{array}{c} A-1 \\ x-1 \end{array}\right).$$

Thus the expression for the expectation of X can be written as

$$E(X) = \frac{A}{\binom{A+B}{n}} \sum_{X'} \binom{A-1}{x-1} \binom{B}{(n-1)-(x-1)},$$

Note that we can "add back in" the value of $x = 0$ if we allow the value $x = 0$ in the summation limits by shifting the indexes "up by one" as

$$\sum_{X'} \binom{A-1}{x-1} \binom{B}{(n-1)-(x-1)} = \sum_{X''} \binom{A-1}{x} \binom{B}{n-1-x}.$$

To simplify this remaining sum we recall **Vandermonde's identity** given by

$$\sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k} = \binom{m+n}{r}. \tag{13}$$

From this we see that the combinatorial sum in $E(X)$ can be simplified to give $\binom{A+B-1}{n-1}$ and the expression we subsequently get for $E(X)$ is the following

$$\begin{aligned} E(X) &= \frac{A}{\binom{A+B}{n}} \binom{A+B-1}{n-1} \\ &= A \cdot \frac{(A+B-1)!}{(n-1)!(A+B-1-n+1)!} \cdot \frac{n!(A+B-n)!}{(A+B)!} \\ &= \frac{An}{A+B}, \end{aligned} \tag{14}$$

as we were to show. Next we evaluate $E(X^2)$. Using the same manipulations as above we compute

$$\begin{aligned} E(X^2) &= \frac{1}{\binom{A+B}{n}} \sum x^2 \binom{A}{x} \binom{B}{n-x} \\ &= \frac{A}{\binom{A+B}{n}} \sum x \binom{A-1}{x-1} \binom{B}{n-x} \\ &= \frac{A}{\binom{A+B}{n}} \left[\sum (x-1) \binom{A-1}{x-1} \binom{B}{n-x} + \sum \binom{A-1}{x-1} \binom{B}{n-x} \right] \\ &= \frac{A}{\binom{A+B}{n}} \left[(A-1) \sum \binom{A-2}{x-2} \binom{B}{n-x} + \binom{A+B-1}{n-1} \right] \\ &= \frac{A(A-1)}{\binom{A+B}{n}} \sum \binom{A-2}{x} \binom{B}{n-2-x} + \frac{An}{A+B} \end{aligned}$$

$$
\begin{aligned}
&= \frac{A(A-1)}{\binom{A+B}{n}} \binom{A+B-2}{n-2} + \frac{An}{A+B} \\
&= A(A-1) \cdot \frac{n!(A+B-n)!}{(A+B)!} \cdot \frac{(A+B-2)!}{(n-2)!(A+B-2-(n-2))!} + \frac{An}{A+B} \\
&= \frac{A(A-1)n(n-1)}{(A+B)(A+B-1)} + \frac{An}{A+B} \,.
\end{aligned}
$$

We can now compute $\text{Var}(X)$ since we have $E(X^2)$ and $E(X)$. We find

$$
\begin{aligned}
\text{Var}(X) &= E(X^2) - E(X)^2 \\
&= \frac{A(A-1)n(n-1)}{(A+B)(A+B-1)} + \frac{An}{A+B} - \frac{A^2n^2}{(A+B)^2} \\
&= \frac{nAB}{(A+B)^2} \cdot \frac{A+B-n}{A+B-1} \,, \qquad (15)
\end{aligned}
$$

when we simplify. This is the result we wanted tho show.

Exercise 15 (the characteristic function of a normal random variable)

For a p.d.f given by $f(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, we will try to evaluate $\zeta(t)$ directly. We have

$$
\zeta(t) = E(e^{itX}) = \frac{1}{(2\pi)^{1/2}\sigma} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \,.
$$

The argument of the exponential in the above expression is given by

$$
\begin{aligned}
& -\frac{1}{2\sigma^2}\left[x^2 - 2\mu x + \mu^2 - 2i\sigma^2 tx\right] \\
&= -\frac{1}{2\sigma^2}\left[x^2 - 2(\mu + i\sigma^2 t)x + \mu^2\right] \\
&= -\frac{1}{2\sigma^2}\left[x^2 - 2(\mu + i\sigma^2 t)x + (\mu + i\sigma^2 t)^2 - (\mu + i\sigma^2 t)^2 + \mu^2\right] \\
&= -\frac{1}{2\sigma^2}(x - (\mu + i\sigma^2 t))^2 + \frac{(\mu + i\sigma^2 t)^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} \\
&= -\frac{1}{2\sigma^2}(x - (\mu + i\sigma^2 t))^2 + \frac{\mu^2 + 2\mu\sigma^2 ti - \sigma^4 t^2 - \mu^2}{2\sigma^2} \\
&= -\frac{1}{2\sigma^2}(x - (\mu + i\sigma^2 t))^2 + \frac{2\mu\sigma^2 ti - \sigma^4 t^2}{2\sigma^2} \,.
\end{aligned}
$$

Thus the integral expression we seek to evaluate looks like

$$
\zeta(t) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{it\mu - \frac{\sigma^2 t^2}{2}\right\} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-(\mu+i\sigma^2 t))^2} dx \,.
$$

To evaluate this let $v = x - (\mu + i\sigma^2 t)$ so that $dx = dv$ and the integral above becomes

$$
\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}v^2} dv = (2\pi)^{1/2}\sigma \,.
$$

Thus the characteristic function for a Gaussian random variable is given by

$$\zeta(t) = \exp\left\{ it\mu - \frac{\sigma^2 t^2}{2} \right\}, \tag{16}$$

as we were to show. We can use this result to compute the expectation of X as follows

$$\begin{aligned}
E(X) &= \frac{1}{i} \frac{\partial \zeta(t)}{\partial t}\bigg|_{t=0} \\
&= \frac{1}{i} \exp\left\{ it\mu - \frac{t^2\sigma^2}{2} \right\} (i\mu - t\sigma^2)\bigg|_{t=0} \\
&= \mu .
\end{aligned}$$

The value of $E(X^2)$ can be computed in the same way

$$\begin{aligned}
E(X^2) &= \frac{1}{i^2} \frac{\partial^2 \zeta(t)}{\partial t^2}\bigg|_{t=0} \\
&= \frac{1}{i^2} \frac{\partial}{\partial t} \left((i\mu - t\sigma^2) \exp\left\{ it\mu - \frac{t^2\sigma^2}{2} \right\} \right)\bigg|_{t=0} \\
&= \frac{1}{i^2} \left[-\sigma^2 \exp\left\{ it\mu - \frac{t^2\sigma^2}{2} \right\} + (i\mu - t\sigma^2)(i\mu - t\sigma^2) \exp\left\{ it\mu - \frac{t^2\sigma^2}{2} \right\} \right]\bigg|_{t=0} \\
&= -\left[-\sigma^2 + (i\mu)^2 \right] \\
&= \sigma^2 + \mu^2 .
\end{aligned}$$

Using this we can compute the variance of X. We find

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \sigma^2 ,$$

as we were to show.

Exercise 16 (the distribution of $\sum_{i=1}^k a_i X_i + b$)

Lets first derive the p.d.f of the random variable $a_i X_i$ when X_i is a normal random variable with a mean given by μ_i and a variance given by σ_i^2. To do this let $Y_i = a_i X_i$ then we find that the expression for $g_Y(y)$ is given by

$$\begin{aligned}
g_Y(y) &= g_X(x(y)) \left| \frac{dx}{dy} \right| = \frac{1}{|a_i|} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{ -\frac{1}{2} \frac{(y_i/a_i - \mu_i)^2}{\sigma_i^2} \right\} \\
&= \frac{1}{\sqrt{2\pi}|a_i|\sigma_i} \exp\left\{ -\frac{1}{2} \frac{(y_i - a_i\mu_i)^2}{a_i^2\sigma_i^2} \right\} ,
\end{aligned}$$

which is the p.d.f of a normal random variable with a mean given by $a_i\mu_i$ and a variance given by $a_i^2\sigma_i^2$. This random variable has a characteristic function given by

$$\zeta_{Y_i}(t) = \exp\left\{ it\mu_i a_i - \frac{t^2 a_i^2 \sigma_i^2}{2} \right\} ,$$

so the sum of k independent random variables has a characteristic function that is the product of these individual characteristic functions as

$$\zeta_{Y_1+Y_2+\cdots+Y_k}(t) = \exp\left\{ it\sum_{i=1}^{k}\mu_i a_i - \frac{t^2}{2}\sum_{i=1}^{k} a_i^2\sigma_i^2 \right\},$$

which is the characteristic function of a Gaussian random variable with a mean $\sum_{i=1}^k \mu_i a_i$ and a variance $\sum_{i=1}^k a_i^2\sigma_i^2$. Lets now define the random variables L and Z to be $Z = \sum_{i=1}^k a_i X_i$ and $L = Z + b$. From the discussion above we know that Z is a Gaussian random variable with a mean given by $\sum_{i=1}^k \mu_i a_i$ and a variance $\sum_{i=1}^k a_i^2\sigma_i^2$. To finish this problem we want to know what the distribution of L is. Consider the distribution function for L. We have

$$F_L(l) = \Pr\{Z + b < l\} = \Pr\{Z < l - b\} = F_Z(l - b),$$

so $f_L(l)$ (the p.d.f. of the random variable L is given by) is obtained by taking the derivative of $F_Z(l-b)$ with respect to l where we find

$$f_L(l) = F_Z'(l-b) = f_Z(l-b) = \frac{1}{(2\pi)^{1/2}\sqrt{\sum_{i=1}^k a_i^2\sigma_i^2}} \exp\left\{ -\frac{1}{2}\frac{(l - b - \sum_{i=1}^k a_i\mu_i)^2}{\sum_{i=1}^k a_i^2\sigma_i^2} \right\}.$$

This later expression is a Gaussian density with mean $b + \sum_{i=1}^k a_i\mu_i$ and variance $\sum_{i=1}^k a_i^2\sigma_i^2$ as we were to show.

Exercise 17 ($X_1 + X_2$ and $X_1 - X_2$ are independent)

Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$ so to work this problem we want to show independence of the two random variables Y_1 and Y_2. Introduce the vectors $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ and the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, so that the vector Y in terms of the vector X is given by $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = AX$. As X_1 and X_2 are independent random draws from the same normal distribution their covariance matrix Σ_X is given by $\Sigma_X = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$. Because Y is a linear transformation of X the covariance matrix for the vector Y is thus given by

$$\begin{aligned}
\Sigma_Y &= A\Sigma_X A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\
&= \sigma^2\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \sigma^2\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2\sigma^2 I.
\end{aligned}$$

From this expression we see that the determinant of Σ_Y is given by $|\Sigma_Y| = (2\sigma^2)^2$. The mean of the vector Y is given by $\begin{bmatrix} 2\mu \\ 0 \end{bmatrix}$ where μ is the mean value of X_i. Given this mean vector and covariance matrix the p.d.f. of the vector random variable Y then takes the form given by

$$g_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)|\Sigma_Y|^{1/2}} \exp\left\{ -\frac{1}{2}\left(\mathbf{y} - \left[\begin{array}{c} 2\mu \\ 0 \end{array}\right]\right)^T (2\sigma^2 I)^{-1} \left(\mathbf{y} - \left[\begin{array}{c} 2\mu \\ 0 \end{array}\right]\right)\right\} .$$

The inner product in the argument to the exponential simplifies

$$\left(\mathbf{y} - \left[\begin{array}{c} 2\mu \\ 0 \end{array}\right]\right)^T (2\sigma^2 I)^{-1} \left(\mathbf{y} - \left[\begin{array}{c} 2\mu \\ 0 \end{array}\right]\right) = \frac{1}{2\sigma^2}\left[(y_1 - 2\mu)^2 + y_2^2\right] ,$$

so $g_{\mathbf{Y}}(\mathbf{y})$ becomes

$$g_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\sqrt{2\pi}\sqrt{2\sigma^2}} \exp\left\{ -\frac{1}{2}\frac{(y_1 - 2\mu)^2}{2\sigma^2}\right\} \cdot \frac{1}{\sqrt{2\pi}\sqrt{2\sigma^2}} \exp\left\{ -\frac{1}{2}\frac{y_2^2}{2\sigma^2}\right\} .$$

Since the random variable Y_1 has a mean of 2μ and a variance of $2\sigma^2$ and the random variable Y_2 has a mean of 0 and a variance of $2\sigma^2$ the above show the joint distribution of (Y_1, Y_2) is the product of two marginal densities showing the independence of Y_1 and Y_2.

Exercise 18 (the gamma function at $1/2$ and other miscellanea)

The second part of this problem is to evaluate $\Gamma(1/2)$. This expression is defined as

$$\Gamma(1/2) = \int_0^\infty x^{-1/2} e^{-x} dx .$$

Since the argument of the exponential is the square of the term $x^{1/2}$ this observation might motivate the substitution $y = \sqrt{x}$. Following the hint let $y = \sqrt{2x}$, so that

$$dy = \frac{1}{\sqrt{2x}} dx .$$

So that with this substitution $\Gamma(1/2)$ becomes

$$\Gamma(1/2) = \int_0^\infty \sqrt{2}\, dy\, e^{-y^2/2} = \sqrt{2}\int_0^\infty e^{-y^2/2} dy .$$

Now from the normalization of the standard Gaussian we know that

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} \exp\{-\frac{y^2}{2}\} dy = 1 ,$$

which easily transforms (by integrating only over the positive real numbers) into

$$2\int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\{-\frac{y^2}{2}\} dy = 1 .$$

Finally manipulating this into the specific integral required to evaluate $\Gamma(1/2)$ we find that

$$\sqrt{2}\int_0^\infty \exp\{-\frac{y^2}{2}\} dy = \sqrt{\pi} ,$$

which shows that $\Gamma(1/2) = \sqrt{\pi}$ as requested.

Exercise 19 (properties of gamma random variables)

If X is given by a gamma distribution then it has a p.d.f given by

$$f(x|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} . \tag{17}$$

The characteristic function for a gamma random variable is then given by

$$\begin{aligned} \zeta(t) &= E(e^{itX}) = \int_{x=0}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{itx} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{x=0}^{\infty} x^{\alpha-1} e^{-(\beta - it)x} dx . \end{aligned}$$

To evaluate this integral let $v = (\beta - it)x$ so that $x = \frac{v}{\beta - it}$ and $dv = (\beta - it)dx$ and we get

$$\zeta(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\beta - it)^{\alpha-1}} \int_{v=0}^{\infty} v^{\alpha-1} e^{-v} \frac{dv}{\beta - it} .$$

If we recall the definition of the Gamma function

$$\Gamma(\alpha) \equiv \int_{v=0}^{\infty} v^{\alpha-1} e^{-v} dv , \tag{18}$$

we see that the above integral becomes

$$\begin{aligned} \zeta(t) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\beta - it)^\alpha} = \left(\frac{\beta}{\beta - it} \right)^\alpha \\ &= \left(1 - \frac{it}{\beta} \right)^{-\alpha} , \end{aligned} \tag{19}$$

as requested. Using this expression we could compute $E(X)$ and $E(X^2)$ via derivatives. Alternatively we could compute these expectations directly as follows

$$\begin{aligned} E(X) &= \int_{x=0}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_{v=0}^{\infty} \frac{v^\alpha}{\beta^\alpha} e^{-v} \frac{dv}{\beta} \\ &= \frac{1}{\beta \Gamma(\alpha)} \int_{v=0}^{\infty} v^\alpha e^{-v} dv = \frac{\Gamma(\alpha+1)}{\beta \Gamma(\alpha)} = \frac{\alpha}{\beta} , \end{aligned} \tag{20}$$

when we make the substitution $v = \beta x$. Next we find $E(X^2)$ given by

$$\begin{aligned} E(X^2) &= \int_{x=0}^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\beta x} dx = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\beta^{\alpha+1}} \frac{1}{\beta} \int_{v=0}^{\infty} v^{\alpha+1} e^{-v} dv \\ &= \frac{1}{\beta^2 \Gamma(\alpha)} \Gamma(\alpha+2) = \frac{(\alpha+1)\alpha}{\beta^2} . \end{aligned} \tag{21}$$

Thus the variance is given by

$$\mathrm{Var}(X) = \frac{(\alpha+1)\alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2} , \tag{22}$$

as we were to show.

Exercise 20 (when X_i is a gamma random variable what is $c(X_1 + X_2 + \cdots X_k)$)

If the independent variables X_i have a gamma distribution with parameters α_i and β we want to show that

$$c(X_1 + X_2 + \cdots + X_k)\,,$$

has a gamma distribution with parameters $\alpha_1+\alpha_2+\cdots+\alpha_k$ and β/c. To do this note that if the random variable X_i are independent gamma distributed with parameters α_i and β then the random variable Y_i defined by $Y_i \equiv cX_i$ has a p.d.f given by

$$\begin{aligned} f_Y(y) &= f_X(x(y))\left|\frac{dx}{dy}\right| = \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)}\left(\frac{y}{c}\right)^{\alpha_i-1} e^{-\beta y/c}\left(\frac{1}{c}\right) \\ &= \frac{(\beta/c)^{\alpha_i}}{\Gamma(\alpha_i)} y^{\alpha_i-1} e^{-(\beta/c)y}\,, \end{aligned}$$

which is a gamma p.d.f. with parameters α_i and β/c. Thus using Equation 19 the characteristic function for this random variable Y_i is given by

$$\zeta_{Y_i}(t) = \left(1 - \frac{it}{(\beta/c)}\right)^{-\alpha_i}\,.$$

Since we want the random variable that is the *sum* of these Y_i the characteristic function for this sum is the product of these individual characteristic functions. Thus

$$\zeta_{Y_1+Y_2+\cdots+Y_k}(t) = \zeta_{c(X_1+X_2+\cdots+X_k)}(t) = \left(1 - \frac{it}{(\beta/c)}\right)^{-\sum_{i=1}^k \alpha_i}\,,$$

which is the characteristic function for a gamma random variable with parameters $\sum_{i=1}^k \alpha_i$ and β/c as we were to show.

Exercise 21 (the p.d.f. of the distribution $\sum_{i=1}^k X_i^2$ when X_i is Gaussian)

For this problem lets define the random variable $Z = \sum_{i=1}^k X_i^2$ and attempt to compute the distribution function for the random variable Z. We have

$$\begin{aligned} F_Z(z) &= \Pr\{Z \le z\} = \Pr\left\{\sum_{i=1}^k X_i^2 \le z\right\} \\ &= \int_{\sum_{i=1}^k x_i^2 \le z} p(x_1, x_2, \cdots, x_k) d\mathbf{x} \\ &= \int_{\sum_{i=1}^k x_i^2 \le z} \prod_{i=1}^k p(x_i) d\mathbf{x} \\ &= \int_{\sum_{i=1}^k x_i^2 \le z} \prod_{i=1}^k \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_i^2\right\} d\mathbf{x} \\ &= \int_{\sum_{i=1}^k x_i^2 \le z} \frac{1}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^k x_i^2\right\} d\mathbf{x}\,. \end{aligned}$$

To evaluate this last integral we will change from Cartesian coordinates to polar coordinates. To do this we recognize that the above integral is an integral over all points $\mathbf{x} \in \mathbb{R}^k$ such that $x^T x < 1$. Let $r^2 = \sum_{i=1}^k x_i^2$ and we get that our differential of volume $d\mathbf{x}$, written in terms of spherical coordinates is given by

$$d\mathbf{x} = kC_k r^{k-1} dr \,, \tag{23}$$

where C_k is the volume of the unit k-sphere and is given by

$$C_k = \frac{\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2}+1)} \,. \tag{24}$$

We require this general expression for $d\mathbf{x}$ since we are working with $\mathbf{x} \in \mathbb{R}^k$ but we can easily verify its correctness by computing $d\mathbf{x}$ in lower dimensions say $k = 2$ (the disk) and $k = 3$ (the sphere). We find

$$\begin{aligned} d\mathbf{x} &= 3r^2 C_3 dr = 3r^2 \left(\frac{4\pi}{3}\right) dr = 4\pi r^2 dr \quad \text{when } k = 3 \quad \text{and} \\ d\mathbf{x} &= 2rC_2 dr = 2\pi r dr \quad \text{when } k = 2 \,. \end{aligned}$$

Using the above general expression for $d\mathbf{x}$ in polar we find that

$$F_Z(z) = \frac{1}{(2\pi)^{k/2}} \int_{r=0}^{\sqrt{z}} e^{-\frac{1}{2}r^2} kC_k r^{k-1} dr = \frac{kC_k}{(2\pi)^{k/2}} \int_{r=0}^{\sqrt{z}} r^{k-1} e^{-\frac{1}{2}r^2} dr \,.$$

As there is no way to evaluate this last integral explicitly we will take the derivative of $F_z(z)$ to get the p.d.f for Z. We find

$$\begin{aligned} f_Z(z) &= F_Z'(z) = \frac{kC_k}{(2\pi)^{k/2}} z^{\frac{k-1}{2}} e^{-\frac{1}{2}z} \left(\frac{1}{2} z^{-1/2}\right) \\ &= \frac{k}{2^{\frac{k}{2}+1}\Gamma(\frac{k}{2}+1)} z^{\frac{k}{2}-1} e^{-\frac{1}{2}z} \,, \end{aligned}$$

when we use Equation 24 for C_k. Since $\Gamma(\frac{k}{2}+1) = \frac{k}{2}\Gamma(\frac{k}{2})$ the above p.d.f. simplifies to

$$f_Z(z) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} z^{\frac{k}{2}-1} e^{-\frac{1}{2}z} \,,$$

which is the p.d.f. of a χ^2 random variable with k degrees of freedom as we were to show.

Exercise 22 (when X_i are exponential R.V.s then $\min(X_i)$ is exponential)

To solve this problem we will use the fact that the random variable that is the minimization of several random variables can be computed by taking minimizations of pairs of random variables. For example if $Z = \min(X_1, X_2, X_3, X_4)$ then the value for Z can be computed as

$$Z = \min(\min(\min(X_1, X_2), X_3), X_4) \,.$$

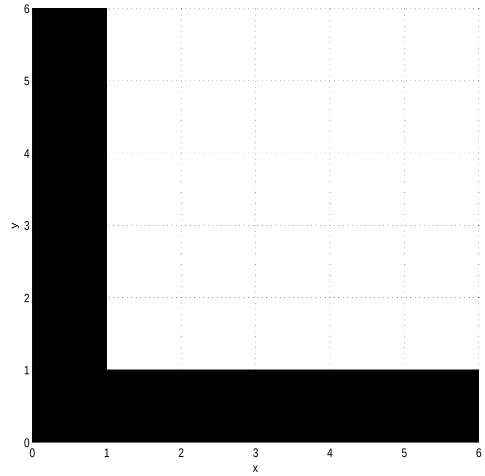

Figure 1: The integration region $\Omega : \{\min(X_1, X_2) \le z\}$, in the (X_1, X_2) plane. Here $z = 1$ for demonstration.

Thus if we can show that the distribution of the random variable $Z = \min(X_1, X_2)$ when X_1 and X_2 are independent exponential random variables with parameters β_1 and β_2 respectively is another exponential random variable with parameters $\beta_1+\beta_2$, then by the nesting property above the distribution of the minimization of k such exponential random variables is another exponential random variable with parameter $\sum_{i=1}^{k} \beta_i$. Thus for the remainder of this problem we show that the distribution of $Z = \min(X_1, X_2)$ is exponential with a parameter $\beta_1 + \beta_2$. Consider the distribution function for the random variable Z. We have that

$$F_Z(z) = \Pr\{\min(X_1, X_2) \le z\} = \int_{\Omega:\{\min(X_1,X_2)\le z\}} p_{X_1}(x_1) p_{X_2}(x_2) dx_1 dx_2 \,.$$

To compute this integral we need to integrate the above expression over the domain shown in Figure 1. This integration region can be represented mathematically as the following two integrals (be careful not to count the region in the lower left corner twice)

$$F_Z(z) = \int_{x_1=0}^{z} \int_{x_2=0}^{\infty} p_{X_1}(x_1) p_{X_2}(x_2) dx_2 dx_1 + \int_{x_2=0}^{z} \int_{x_1=z}^{\infty} p_{X_1}(x_1) p_{X_2}(x_2) dx_1 dx_2 \,.$$

When we put in the expressions for the exponential distributions we get.

$$F_Z(z) = \beta_1\beta_2 \int_{x_1=0}^{z} e^{-\beta_1 x_1} \int_{x_2=0}^{\infty} e^{-\beta_2 x_2} dx_2 dx_1 + \beta_1\beta_2 \int_{x_2=0}^{z} e^{-\beta_2 x_2} \int_{x_1=z}^{\infty} e^{-\beta_1 x_1} dx_1 dx_2 \,.$$

Plots of the integration region are performed in the Matlab file `chap_4_prob_22.m`. We will evaluate the first integral. We find that

$$\begin{aligned}
\beta_1\beta_2 \int_{x_1=0}^{z} e^{-\beta_1 x_1} \int_{x_2=0}^{\infty} e^{-\beta_2 x_2} dx_2 dx_1 &= \beta_1\beta_2 \int_{x_1=0}^{z} e^{-\beta_1 x_1} \left(\left. \frac{e^{-\beta_2 x_2}}{-\beta_2} \right|_{x_2=0}^{\infty} \right) dx_1 \\
&= \beta_1 \int_{x_1=0}^{z} e^{-\beta_1 x_1} dx_1 = 1 - e^{-\beta_1 z} \,.
\end{aligned}$$

Next we evaluate the second integral. We have

$$\beta_1\beta_2 \int_{x_2=0}^{z} e^{-\beta_2 x_2} \int_{x_1=z}^{\infty} e^{-\beta_1 x_1} dx_1 dx_2 = \beta_1\beta_2 \int_{x_2=0}^{z} e^{-\beta_2 x_2} \left(\left. \frac{e^{-\beta_1 x_1}}{-\beta_1} \right|_{x_1=z}^{\infty} \right) dx_2$$

$$\begin{aligned}
&= \beta_2 e^{-\beta_1 z} \int_{x_2=0}^{z} e^{-\beta_2 x_2} dx_2 \\
&= e^{-\beta_1 z} \left(1 - e^{-\beta_2 x_2}\right) .
\end{aligned}$$

when we combine these two integrals (by adding we find)

$$F_Z(z) = 1 - e^{-(\beta_1+\beta_2)z} ,$$

this is the distribution function for an exponential random variable with a parameter $\beta_1 + \beta_2$ as we were to show.

Exercise 23 (the distribution of the first differences)

For this problem we are told that X_i is given by an exponential distribution and thus has a p.d.f. given by $\beta e^{-\beta x}$ and Y_i are the order statistics of n draws of the random variables X_i for $i = 1, 2, \cdots, n$. Introduce the random variables Z_i defined as

$$\begin{aligned}
Z_1 &= nY_1 \\
Z_2 &= (n-1)(Y_2 - Y_1) \\
Z_3 &= (n-2)(Y_3 - Y_2) \\
&\vdots \\
Z_{n-1} &= 2(Y_{n-1} - Y_{n-2}) \\
Z_n &= (Y_n - Y_{n-1}) .
\end{aligned}$$

Then to get the p.d.f. of the vector $\mathbf{Z}$ defined in this way we recall that

$$g_{Z_1,Z_2,\cdots,Z_n}(z_1, z_2, \cdots, z_n) = f_{Y_1,Y_2,\cdots,Y_n}(y_1, y_2, \cdots, y_n) \left| \frac{\partial \mathbf{Y}}{\partial \mathbf{Z}} \right| .$$

From the above specified relationship between $\mathbf{Y}$ and $\mathbf{Z}$ we have

$$\frac{\partial \mathbf{Z}}{\partial \mathbf{Y}} = \left[\begin{array}{ccccccc}
n & 0 & & & & & \\
-(n-1) & n-1 & 0 & & & & \\
0 & -(n-2) & n-2 & & & & \\
 & 0 & -(n-3) & \ddots & & & \\
 & & & \ddots & 3 & 0 & \\
 & & & & -2 & 2 & 0 \\
 & & & & 0 & -1 & 1
\end{array} \right] . \tag{25}$$

Thus the determinant of the above matrix is given by

$$\left| \frac{\partial \mathbf{Z}}{\partial \mathbf{Y}} \right| = n(n-1)(n-2)(n-3)\cdots 321 = n! ,$$

so

$$\left| \frac{\partial \mathbf{Y}}{\partial \mathbf{Z}} \right| = \frac{1}{n!} .$$

Then as we know the distribution of the order statistics as

$$\begin{aligned} f_{Y_1,Y_2,\cdots,Y_n}(y_1(\mathbf{z}), y_2(\mathbf{z}), \cdots, y_n(\mathbf{z})) &= n! f(y_1) f(y_2) \cdots f(y_n) \\ &= n! \prod_{i=1}^{n} \beta e^{-\beta y_i} = n! \beta^n e^{-\beta \sum_{i=1}^{n} y_i} . \end{aligned}$$

Now we will compute $\sum_{i=1}^{n} y_i$ in terms of the elements z_i. Consider the following inner product expression for $\sum_{i=1}^{n} z_i$

$$\begin{aligned} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} &= \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} n & 0 & & & & & \\ -(n-1) & n-1 & 0 & & & & \\ 0 & -(n-2) & n-2 & & & & \\ & 0 & -(n-3) & \ddots & & & \\ & & & \ddots & 3 & 0 & \\ & & & & -2 & 2 & 0 \\ & & & & 0 & -1 & 1 \end{bmatrix} \\ &\times \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} , \end{aligned}$$

when we use the relationship between z_i and y_i and compute the product of the vector $\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$ with the lower diagonal matrix given by Equation 25. Thus from this result we see that $\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} y_i$, thus we find that our p.d.f. of $\mathbf{Z}$ is given by

$$g_{Z_1,Z_2,\cdots,Z_n}(z_1, z_2, \cdots, z_n) = n! \beta^n e^{-\beta \sum_{i=1}^{n} z_i} \frac{1}{n!} = \beta^n e^{-\beta \sum_{i=1}^{n} z_i} ,$$

or the p.d.f. of a set of n independent exponential random variables which is the same as the joint distribution of $X_1, X_2, \cdots X_n$ and what we were to show.

Exercise 24 (from the conditional and marginal get the other marginal)

We are told the conditional p.d.f. for the random variables X given Y is a Poisson random variable and the marginal p.d.f. for the variable Y is a gamma random variable with parameters α and β. Thus they are expressed as

$$\begin{aligned} P(X = x | Y = y) &= \frac{e^{-y} y^x}{x!} \\ P(Y = y) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha - 1} e^{-\beta y} . \end{aligned}$$

We can evaluate the marginal distribution of X or $P(X = x)$ by conditioning on the value of Y as follows

$$P(X = x) = \int P(X = x | Y = y) P(Y = y) dy$$

$$
\begin{aligned}
&= \int_0^\infty \frac{e^{-y}y^x}{x!}\frac{\beta^\alpha}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}dy \\
&= \frac{\beta^\alpha}{x!\Gamma(\alpha)}\int_0^\infty y^{x+\alpha-1}e^{-(1+\beta)y}dy\,.
\end{aligned}
$$

To evaluate this integral let $v = (1+\beta)y$ so that $dv = (1+\beta)dy$ and $dy = \frac{dv}{1+\beta}$ and our integral becomes

$$
\begin{aligned}
P(X=x) &= \frac{\beta^\alpha}{x!\Gamma(\alpha)}\int_{v=0}^\infty \frac{v^{x+\alpha-1}}{(1+\beta)^{x+\alpha-1}}e^{-v}\frac{dv}{1+\beta} \\
&= \frac{\beta^\alpha}{x!\Gamma(\alpha)(1+\beta)^{x+\alpha}}\int_0^\infty v^{x+\alpha-1}e^{-v}dv \\
&= \frac{\beta^\alpha\Gamma(x+\alpha)}{x!\Gamma(\alpha)(1+\beta)^{x+\alpha}}\,.
\end{aligned}
$$

If this expression is to equal a negative binomial random variable with parameters r and p it must have a form given by

$$
\left(\begin{array}{c} r+x-1 \\ x \end{array}\right) p^r q^x\,.
$$

To write the above in this form note that since $\Gamma(n) = (n-1)!$ when n is an positive integer the ratio of gamma functions in the above expression

$$
\frac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)} = \frac{(x+\alpha-1)!}{x!(\alpha-1)!} = \left(\begin{array}{c} x+\alpha-1 \\ x \end{array}\right).
$$

Thus using this we have $P(X=x)$ given by

$$
P(X=x) = \left(\begin{array}{c} x+\alpha-1 \\ x \end{array}\right)\frac{\beta^\alpha}{(1+\beta)^{x+\alpha}}\,.
$$

To make this match the negative binomial expression we can take $r = \alpha$ and $p = \frac{\beta}{1+\beta}$), so that $q = 1-p = \frac{1}{1+\beta}$, and showing the desired equivalence.

Exercise 25 (the normalization of the beta p.d.f.)

For this problem we want to show that

$$
\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx\,, \tag{26}
$$

where the Gamma function is defined by

$$
\Gamma(\alpha) = \int_0^\infty v^{\alpha-1}e^{-v}dv\,. \tag{27}
$$

With this definition we see that the product of $\Gamma(\alpha)$ and $\Gamma(\beta)$ can be written as a double integral as

$$
\Gamma(\alpha)\Gamma(\beta) = \int_{r=0}^\infty\int_{s=0}^\infty s^{\alpha-1}r^{\beta-1}e^{-(r+s)}dsdr\,.
$$

Lets change the integration variables from s and r to u and v defined by $u = r$ and $v = \frac{s}{r}$. The inverse transformation for this is $r = u$ and $s = rv = uv$. The differential change in volume then transforms as

$$dsdr = \left|\frac{d(r,s)}{d(u,v)}\right| dvdu = \left|\begin{array}{cc} \frac{\partial r}{\partial u} & \frac{\partial r}{\partial v} \\ \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \end{array}\right| dvdu = \left|\begin{array}{cc} 1 & 0 \\ v & u \end{array}\right| dvdu = u\, dvdu\,.$$

Using this change of coordinates we get that the product of $\Gamma(\alpha)\Gamma(\beta)$ is given by

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_{u=0}^{\infty}\int_{v=0}^{\infty} u^{\alpha-1}v^{\alpha-1}u^{\beta-1}e^{-u(1+v)}udvdu \\
&= \int_{u=0}^{\infty}\int_{v=0}^{\infty} u^{\alpha+\beta-1}v^{\alpha-1}e^{-(1+v)u}dvdu \\
&= \int_{v=0}^{\infty} v^{\alpha-1}\int_{u=0}^{\infty} u^{\alpha+\beta-1}e^{-(1+v)u}dudv\,.
\end{aligned}$$

Lets evaluate the inner integral over u. To do this let $\xi = (1+v)u$ so that $d\xi = (1+v)du$ and we have

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_{v=0}^{\infty} v^{\alpha-1}\int_{\xi=0}^{\infty} \frac{\xi^{\alpha+\beta-1}}{(1+v)^{\alpha+\beta-1}}e^{-\xi}\left(\frac{d\xi}{1+v}\right)dv \\
&= \int_{v=0}^{\infty} \frac{v^{\alpha-1}}{(1+v)^{\alpha+\beta}}\int_{\xi=0}^{\infty} \xi^{\alpha+\beta-1}e^{-\xi}d\xi dv \\
&= \Gamma(\alpha+\beta)\int_{v=0}^{\infty}\frac{v^{\alpha-1}}{(1+v)^{\alpha+\beta}}dv\,.
\end{aligned}$$

Next in the remaining integral let $u = \frac{1}{1+v}$ so that $v = \frac{1}{u} - 1 = \frac{1-u}{u}$ and $du = -\frac{1}{(1+v)^2}dv = -u^2dv$ and the above becomes

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \Gamma(\alpha+\beta)\int_{u=1}^{0}\left(\frac{1-u}{u}\right)^{\alpha-1}u^{\alpha+\beta}\left(-\frac{du}{u^2}\right) \\
&= \Gamma(\alpha+\beta)\int_{u=0}^{1}(1-u)^{\alpha-1}u^{\alpha+\beta-(\alpha-1)-2}du \\
&= \Gamma(\alpha+\beta)\int_{u=0}^{1}(1-u)^{\alpha-1}u^{\beta-1}du\,,
\end{aligned}$$

as we were to show.

Exercise 26 (statistics of the beta distribution)

To begin we attempt to evaluate $E(X^m(1-X)^n)$, when X is given by a beta random variable. We find that recalling the normalization of the beta function that

$$\begin{aligned}
E(X^m(1-X)^n) &= \int x^m(1-x)^n\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^1 x^{m+\alpha-1}(1-x)^{n+\beta-1}dx
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(m+\alpha)\Gamma(n+\beta)}{\Gamma(m+n+\alpha+\beta)} \\
&= \frac{\Gamma(m+\alpha)}{\Gamma(\alpha)} \cdot \frac{\Gamma(n+\beta)}{\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(m+n+\alpha+\beta)} \,. \qquad (28)
\end{aligned}
$$

Now note that we can simplify some of these ratios of gamma functions as follows

$$
\begin{aligned}
\frac{\Gamma(m+\alpha)}{\Gamma(\alpha)} &= \frac{(m-1+\alpha)\Gamma(m-1+\alpha)}{\Gamma(\alpha)} \\
&= \frac{(m-1+\alpha)(m-2+\alpha)\Gamma(m-2+\alpha)}{\Gamma(\alpha)} \\
&= \frac{(m-1+\alpha)(m-2+\alpha)\cdots(m-(m-1)+\alpha)(m-m+\alpha)\Gamma(\alpha)}{\Gamma(\alpha)} \\
&= (m-1+\alpha)(m-2+\alpha)\cdots(2+\alpha)(1+\alpha)\alpha \\
&= \prod_{k=1}^{m}(m-k+\alpha)\,.
\end{aligned}
$$

In the same way

$$
\frac{\Gamma(n+\beta)}{\Gamma(\beta)} = \prod_{k=1}^{n}(n-k+\beta)\,.
$$

and

$$
\frac{\Gamma(m+n+\alpha+\beta)}{\Gamma(\alpha+\beta)} = \prod_{k=1}^{m+n}(m+n-k+\alpha+\beta)\,.
$$

Using these expressions we find that $E(X^m(1-X)^n)$ is given by

$$
E(X^m(1-X)^n) = \frac{\prod_{i=1}^{m}(m-i+\alpha)\prod_{j=1}^{n}(n-j+\beta)}{\prod_{k=1}^{m+n}(m+n-k+\alpha+\beta)}\,. \qquad (29)
$$

Lets now evaluate this expression for some special values of m and n. If $m=1$ and $n=0$ then we find that

$$
E(X) = \frac{\alpha}{\alpha+\beta}\,, \qquad (30)
$$

If $m=2$ and $n=0$ then we find that

$$
E(X^2) = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}\,, \qquad (31)
$$

so that the variance of a beta distributed random variable is given by

$$
\begin{aligned}
\mathrm{Var}(X) &= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\,. \qquad (32)
\end{aligned}
$$

Exercise 27 (show that $X_1/(X_1+X_2)$ and X_1+X_2 are independent)

For this problem we are told that X_1 has a gamma distribution with parameters α_1 and β, while X_2 also has a gamma distribution with parameters α_2 and β. Introduce the two random variables R and S defined by

$$\begin{aligned} R &= \frac{X_1}{X_1+X_2} \\ S &= X_1+X_2 \,, \end{aligned}$$

Since X_1 and X_2 are independent the joint p.d.f over X_1, X_2 is the product of the two marginals. Thus

$$p_{(X_1,X_2)}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2) = \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} x_1^{\alpha_1-1} e^{-\beta x_1} \cdot \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} x_2^{\alpha_2-1} e^{-\beta x_2} \,.$$

Lets transform the p.d.f. over the vector (X_1, X_2) into a p.d.f over the vector (R, S). The change of variables formula requires that

$$p_{(R,S)}(r,s) = p_{(X_1,X_2)}(x_1(r,s), x_2(r,s)) \left| \frac{\partial(X_1,X_2)}{\partial(R,S)} \right|$$

To evaluate this lets begin by computing $\frac{\partial(R,S)}{\partial(X_1,X_2)}$. We have

$$\frac{\partial(R,S)}{\partial(X_1,X_2)} = \begin{bmatrix} \frac{\partial R}{\partial X_1} & \frac{\partial R}{\partial X_2} \\ \frac{\partial S}{\partial X_1} & \frac{\partial S}{\partial X_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{X_1+X_2} - \frac{X_1}{(X_1+X_2)^2} & -\frac{X_1}{(X_1+X_2)^2} \\ 1 & 1 \end{bmatrix} .$$

Thus we find the determinant of the above expression or $\left| \frac{\partial(R,S)}{\partial(X_1,X_2)} \right|$ given by

$$\left| \frac{\partial(R,S)}{\partial(X_1,X_2)} \right| = \frac{1}{(X_1+X_2)} - \frac{X_1}{(X_1+X_2)^2} + \frac{X_1}{(X_1+X_2)^2} = \frac{1}{X_1+X_2} \,,$$

so the inverse of this expression or $\left| \frac{\partial(X_1,X_2)}{\partial(R,S)} \right|$ is given by

$$\left| \frac{\partial(X_1,X_2)}{\partial(R,S)} \right| = X_1 + X_2 = S \,.$$

Next we solve for X_1 and X_2 in terms of R and S. We find

$$\begin{aligned} X_1 &= RS \\ X_2 &= S - X_1 = S - RS \,, \end{aligned}$$

Thus using these sub-results into the expression for $p_{(R,S)}(r,s)$ we find

$$\begin{aligned} p_{(R,S)}(r,s) &= \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} (rs)^{\alpha_1-1} e^{-\beta rs} \cdot \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} (s-rs)^{\alpha_2-1} e^{-\beta(s-rs)} \, s \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\beta s} s^{\alpha_1+\alpha_2-1} r^{\alpha_1-1} (1-r)^{\alpha_2-1} \,, \end{aligned}$$

Note that if we write this expression as a product with two factors

$$p_{(R,S)}(r,s) = \left(\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} r^{\alpha_1-1}(1-r)^{\alpha_2-1}\right)\left(\frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\beta s} s^{\alpha_1+\alpha_2-1}\right)$$

we see that it is the product of two densities, one over $f_R(r)$ and one over $f_S(s)$. The density over R or $f_R(r) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} r^{\alpha_1-1}(1-r)^{\alpha_2-1}$, is the density of a beta distributed random variable with parameters α_1 and α_2, while the second density over S or $f_S(s) = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\beta s} s^{\alpha_1+\alpha_2-1}$, is the density of a gamma distribution with parameters $\alpha_1+\alpha_2$ and β.

Exercise 28 (statistics of the uniform distribution)

The uniform distribution has a characteristic function that can be computed directly

$$\begin{aligned}\zeta(t) &= E(e^{itX}) = \int_\alpha^\beta e^{itx}\frac{1}{\beta-\alpha}dx \\ &= \frac{1}{\beta-\alpha}\left(\frac{e^{it\beta}-e^{it\alpha}}{it}\right).\end{aligned}$$

We could compute $E(X)$ using the characteristic function $\zeta(t)$ for a uniform random variable. Beginning this calculation we have

$$\begin{aligned}E(X) &= \left.\frac{1}{i}\frac{\partial\zeta(t)}{\partial t}\right|_{t=0} \\ &= \frac{1}{i}\frac{1}{\beta-\alpha}\left.\left[\frac{1}{it}(i\beta e^{it\beta}-i\alpha e^{it\alpha}) - \frac{1}{it^2}(e^{it\beta}-e^{it\alpha})\right]\right|_{t=0} \\ &= -\frac{1}{\beta-\alpha}\left.\left[\frac{t(i\beta e^{it\beta}-i\alpha e^{it\alpha})-(e^{it\beta}-e^{it\alpha})}{t^2}\right]\right|_{t=0}.\end{aligned}$$

To evaluate this expression requires the use of L'Hopital's rule, and seems a somewhat complicated route to compute $E(X)$. The evaluation of $E(X^2)$ would probably be even more work when computed from the characteristic function. For this distribution, it is much easier to compute the expectations directly. We have

$$E(X) = \int_\alpha^\beta x\frac{1}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left.\frac{x^2}{2}\right|_\alpha^\beta = \frac{1}{2}(\alpha+\beta).$$

In the same way we find $E(X^2)$ to be given by

$$\begin{aligned}E(X^2) &= \int_\alpha^\beta x^2\frac{1}{\beta-\alpha}dx = \frac{1}{\beta-\alpha}\left(\frac{\beta^3-\alpha^3}{3}\right) \\ &= \frac{(\beta-\alpha)(\beta^2+\alpha\beta+\alpha^2)}{3(\beta-\alpha)} = \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2).\end{aligned}$$

Using these two results we thus have that the variance of a uniform random variable is

$$\begin{aligned}\text{Var}(X) &= E(X^2)-E(X)^2 \\ &= \frac{1}{3}(\beta^2+\alpha\beta+\alpha^2) - \frac{1}{4}(\alpha^2+\beta^2+2\alpha\beta) \\ &= \frac{(\beta-\alpha)^2}{12}.\end{aligned}$$

Exercise 29 (the joint distribution of the min and the max)

To evaluate the density $f(y,z)$ we recognize that this expression means that after n draws of the variables X_i the max is the value z and the minimum is the value y. Since each of these draws is an independent uniform random variable, unconditionally each X_i has a density given by $\frac{1}{\beta-\alpha}$. Now since once the maximum value of z and the minimum value of y are *specified* all the other $n-2$ draws must have values that fall between y and z. This later event happens with a probability of $\frac{z-y}{\beta-\alpha}$. So for all n draws we would have a probability density proportional to

$$\frac{(z-y)^{n-2}}{(\beta-\alpha)^{n-2}} \cdot \frac{1}{(\beta-\alpha)^2} = \frac{(z-y)^{n-2}}{(\beta-\alpha)^n}.$$

The above expression does not account for the fact that the draws that result from the maximum value z (say) can occur on any of the n draws, while once the maximum is specified the value minimum, z, can occur in any of the $n-1$ draws. Multiplying the above function by these two factors then gives

$$f(y,z) = \frac{n(n-1)(z-y)^{n-2}}{(\beta-\alpha)^n},$$

as requested.

Exercise 30 (moments of the univariate Pareto distribution)

Since the p.d.f for a univariate Pareto distribution is given by

$$f(x|x_0,\alpha) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}, \tag{33}$$

when $x > x_0$ and is 0 otherwise. Note that $E(X^k)$ will only exist the following integral

$$\int_{x_0}^\infty \frac{x^k}{x^{\alpha+1}} dx = \int_{x_0}^\infty x^{k-\alpha-1} dx,$$

converges. The convergence of this later integral requires $k-\alpha-1 < -1$ or

$$k < \alpha.$$

We can compute various statistics for this distribution. We find

$$\begin{aligned} E(X) &= \int_{x_0}^\infty x \frac{\alpha x_0^\alpha}{x^{\alpha+1}} dx = \int_{x_0}^\infty \frac{\alpha x_0^\alpha}{x^\alpha} dx = \alpha x_0^\alpha \left(\left. \frac{x^{-\alpha+1}}{-\alpha+1} \right|_{x_0}^\infty \right) = \alpha x_0^\alpha \left(0 + \frac{x_0^{-\alpha+1}}{\alpha-1} \right) \\ &= \frac{\alpha x_0}{\alpha-1}. \end{aligned} \tag{34}$$

Next assuming $\alpha > 2$ we compute $E(X^2)$. We find that

$$\begin{aligned} E(X^2) &= \int_{x_0}^\infty \frac{\alpha x_0^\alpha}{x^{\alpha-1}} dx = \alpha x_0^\alpha \left(\left. \frac{x^{-\alpha+1+1}}{-\alpha+2} \right|_{x_0}^\infty \right) = \alpha x_0^\alpha \left(0 + \frac{x_0^{-\alpha+2}}{\alpha-2} \right) \\ &= \frac{\alpha x_0^2}{\alpha-2}. \end{aligned} \tag{35}$$

Thus from these two expressions we can compute the variance of X and find

$$\begin{aligned}\text{Var}(X) &= E(X^2) - E(X)^2 = \frac{\alpha x_0^2}{\alpha - 2} - \frac{\alpha^2 x_0^2}{(\alpha-1)^2} \\ &= \frac{\alpha x_0^2}{(\alpha-1)^2(\alpha-2)}\,,\end{aligned} \tag{36}$$

as we were to show.

Exercise 31 (the $\log(X/x_0)$ transformation of a Pareto random variable)

Introduce the random variable Y defined as $Y = \log(X/x_0) = \log(X) - \log(x_0)$, and X is a Pareto distributed random variable. Then the p.d.f. of Y is given by the standard formula for transforming random variables

$$g_Y(y) = g_X(x(y))\left|\frac{dx}{dy}\right|\,,$$

and thus we need to be able to evaluate $\left|\frac{dx}{dy}\right|$. Using the above transformation we see that $X = x_0 e^Y$ and find derivatives of this transformation given by

$$\frac{dy}{dx} = \frac{1}{x} \quad \text{and} \quad \frac{dx}{dy} = x_0 e^y\,.$$

Using this result and the p.d.f of a Pareto distribution we find $g_Y(y)$ given by

$$\begin{aligned}g_Y(y) &= g_X(x(y))|x_0 e^y| \\ &= \frac{\alpha x_0^\alpha}{x_0^{\alpha+1} e^{(\alpha+1)y}} x_0 e^y = \alpha e^{-\alpha y}\,,\end{aligned}$$

which is the p.d.f. of an exponential random variable as we were to show.

Exercise 32 (showing that the ratio $X/(Y/n)^{1/2}$ is a t random variable)

For this problem we are told that X is a random variable distributed as a standard normal and Y is a random variable distributed as a χ^2 random variable with n degrees of freedom. Lets begin by attempting to determine the distribution function for the random variable V defined as

$$V = \frac{X}{\left(\frac{Y}{n}\right)^{1/2}}\,.$$

From the definition of the distribution function and the p.d.f.'s of the random variables X and Y we find

$$F_V(v) = \Pr\left\{\frac{n^{1/2}X}{Y^{1/2}} \le v\right\}$$

$$
\begin{aligned}
&= \Pr\left\{ X \le \frac{Y^{1/2}}{n^{1/2}} v \right\} \\
&= \int_{y=0}^{\infty} \int_{x=-\infty}^{\frac{v}{n^{1/2}} y^{1/2}} p_{X,Y}(x,y) dx dy \\
&= \int_{y=0}^{\infty} \int_{x=-\infty}^{\frac{v}{n^{1/2}} y^{1/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{y^{\frac{n}{2}-1}}{\Gamma(n/2) 2^{n/2}} e^{-y/2} dx dy \\
&= \frac{1}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2}} \int_{y=0}^{\infty} y^{\frac{n}{2}-1} e^{-y/2} \int_{x=-\infty}^{\frac{v}{n^{1/2}} y^{1/2}} e^{-\frac{x^2}{2}} dx dy \,.
\end{aligned}
$$

We can determine the distribution of the random variable V if we have its p.d.f. Since we cannot evaluate the x integral in the above in closed form, lets take the derivative with respect to v and see if we can evaluate the resulting integral for $f_V(v) = F_V'(v)$. We find

$$
\begin{aligned}
f_V(v) &= \frac{1}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2}} \int_{y=0}^{\infty} y^{\frac{n}{2}-1} e^{-y/2} e^{-\frac{yv^2}{2n}} \left(\frac{y^{1/2}}{n^{1/2}} \right) dy \\
&= \frac{1}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2} n^{1/2}} \int_{y=0}^{\infty} y^{\frac{n}{2}-1} \exp\left\{ -\frac{1}{2}\left(1 + \frac{v^2}{n}\right) y \right\} dy \,.
\end{aligned}
$$

To evaluate this let $s = \frac{1}{2}\left(1 + \frac{v^2}{n}\right) y$ so that $ds = \frac{1}{2}\left(1 + \frac{v^2}{n}\right) dy$ and $y = \frac{2s}{1+\frac{v^2}{n}}$ to get

$$
\begin{aligned}
f_V(v) &= \frac{1}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2} n^{1/2}} \int_{s=0}^{\infty} \frac{2^{n/2-1/2} s^{n/2-1/2}}{\left(1 + \frac{v^2}{n}\right)^{\frac{1}{2}(n-1)}} e^{-s} \frac{2ds}{\left(1 + \frac{v^2}{n}\right)} \\
&= \frac{1}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2} n^{1/2}} \frac{1}{\left(1 + \frac{v^2}{n}\right)^{\frac{1}{2}(n+1)}} \int_{s=0}^{\infty} s^{\frac{n}{2}+\frac{1}{2}-1} e^{-s} ds \\
&= \frac{\Gamma(\frac{n}{2} + \frac{1}{2})}{\Gamma(n/2)\pi^{1/2} 2^{n/2+1/2} n^{1/2}} \frac{1}{\left(1 + \frac{v^2}{n}\right)^{\frac{1}{2}(n+1)}} \,,
\end{aligned}
$$

which is the p.d.f of a t random variable with n degrees of freedom as we were to show.

Exercise 33 (statistics of t distributed random variables)

When X has a t distribution with α degrees of freedom it has a p.d.f. that looks like

$$
f(x|\alpha) = \frac{\Gamma((\alpha+1)/2)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)} \left(1 + \frac{x^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)} \,.
$$

Then from this expression the k moment $E(X^k)$ will exist if the following integral

$$
\int_{x=-\infty}^{\infty} \frac{x^k}{\left(1 + \frac{x^2}{\alpha}\right)^{\frac{\alpha+1}{2}}} dx \,,
$$

converges. This later integral converges if the limiting form of $\frac{x^k}{\left(1+\frac{x^2}{\alpha}\right)^{\frac{\alpha+1}{2}}}$ as x goes to infinity or

$$
\frac{x^k}{x^{\alpha+1}} = x^{k-\alpha-1} \,,
$$

has an exponent on x that is smaller than -1, i.e. the function $\frac{1}{x^{\alpha-1-k}}$ must be smaller than $1/x$ or

$$\frac{1}{x^{\alpha-1-k}} < \frac{1}{x}\,.$$

This translates to $k-\alpha-1<-1$ or $k<\alpha$.

The expectation of X or $E(X)$ is given by

$$E(X) = \frac{\Gamma((\alpha+1)/2)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)}\int_{-\infty}^{\infty} x\left(1+\frac{x^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)} dx = 0\,,$$

since this is the integral of an odd function x over a symmetric range.

Next we compute $E(X^2)$. We find the expression for $E(X^2)$ is proportional to the following integral

$$E(X^2) \propto \int_{-\infty}^{\infty} x^2\left(1+\frac{x^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)} dx = 2\int_{0}^{\infty} x^2\left(1+\frac{x^2}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)} dx\,.$$

To evaluate this integral lets change variables to y such that

$$y = \frac{x^2/\alpha}{1+x^2/\alpha} = 1-\frac{1}{1+x^2/\alpha}\,, \tag{37}$$

so that with this definition the common expression $1+\frac{x^2}{\alpha}$ in terms of the variable y is given by

$$1+\frac{x^2}{\alpha} = \frac{1}{1-y}\,.$$

Solving for x^2 in the above transformation we find

$$x^2 = \alpha\frac{y}{1-y} \quad \text{so} \quad x = \sqrt{\alpha}\sqrt{\frac{y}{1-y}}\,.$$

Finally the differential dy in terms of dx using Equation 37 is

$$dy = \frac{2}{(1+x^2/\alpha)^2}\left(\frac{x}{\alpha}\right)dx\,.$$

So dx written in terms of only the variable y is given by

$$dx = \frac{\alpha}{2}\frac{1}{x}\left(1+\frac{x^2}{\alpha}\right)^2 dy = \frac{\sqrt{\alpha}}{2}\frac{1}{\sqrt{y}}\frac{1}{(1-y)^{3/2}}dy\,.$$

Using these expressions we find that $E(X^2)$ proportional to

$$\begin{aligned}
E(X^2) &\propto 2\int_{y=0}^{1}\left(\alpha\frac{y}{1-y}\right)^{\frac{\alpha+1}{2}}\frac{\sqrt{\alpha}}{2}\frac{1}{\sqrt{y}}\frac{1}{(1-y)^{3/2}}dy \\
&= \alpha^{3/2}\int_{y=0}^{1}\sqrt{y}(1-y)^{\frac{\alpha}{2}-2}dy \\
&= \alpha^{3/2}\int_{y=0}^{1}y^{3/2-1}(1-y)^{\frac{\alpha}{2}-1-1}dy \\
&= \frac{\alpha^{3/2}\Gamma(3/2)\Gamma(\frac{\alpha}{2}-1)}{\Gamma(\alpha/2+1/2)}\,.
\end{aligned}$$

Putting back in the proportionality constant we see that $E(X^2)$ is given by

$$E(X^2) = \frac{\Gamma((\alpha+1)/2)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)} \cdot \frac{\alpha^{3/2}\Gamma(3/2)\Gamma(\frac{\alpha}{2}-1)}{\Gamma(\alpha/2+1/2)},$$

or using the fact that $\Gamma(\alpha/2) = (\alpha/2-1)\Gamma(\alpha/2-1)$ and $\Gamma(3/2) = 1/2\Gamma(1/2) = \frac{\sqrt{\pi}}{2}$ we have

$$E(X^2) = \frac{\alpha}{2\left(\frac{\alpha}{2}-1\right)} = \frac{\alpha}{\alpha-2}. \tag{38}$$

Thus since $E(X) = 0$ for a t-distributed random variable the expression for $E(X^2)$ also equals the variance.

Exercise 34 (the P.D.F. of the ratio of normals is a Cauchy distribution)

As stated in the problem, let X_1 and X_2 be distributed as standard normal random variables (i.e. they have mean 0 and variance 1). Then we want the distribution of the variable X_1/X_2. To this end define the random variables U and V as $U = X_1/X_2$ and $V = X_2$. The distribution function of U is then what we are after. From the definition of U and V in terms of X_1 and X_2 we see that $X_1 = UV$ and $X_2 = V$. To solve this problem we will derive the joint distribution function for U and V and then marginalize out V giving the distribution function for U, alone. Now from Theorem $2-4$ on page 45 of Schaum's probability and statistics outline the distribution of the joint random variable (U, V), in term of the joint random variable (X_1, X_2) is given by

$$g(u,v) = f(x_1,x_2)\left|\frac{\partial(x_1,x_2)}{\partial(u,v)}\right|.$$

Now

$$\left|\frac{\partial(x_1,x_2)}{\partial(u,v)}\right| = \left|\begin{array}{cc} v & u \\ 0 & 1 \end{array}\right| = |v|,$$

so that

$$g(u,v) = f(x_1,x_2)|v| = p(x_1)\,p(x_2)|x_2|,$$

as $f(x_1,x_2) = p(x_1)p(x_2)$ since X_1 and X_2 are assumed independent. Now using the fact that the distribution of X_1 and X_2 are standard normals we get

$$g(u,v) = \frac{1}{2\pi}\exp(-\frac{1}{2}(u\,v)^2)\,\exp(-\frac{1}{2}v^2)\,|v|.$$

Marginalizing out the variable V we get

$$g(u) = \int_{-\infty}^{\infty} g(u,v)dv = \frac{1}{\pi}\int_0^{\infty} v\,e^{-\frac{1}{2}(1+u^2)v^2}\,dv.$$

To evaluate this integral let $\eta = \sqrt{\frac{1+u^2}{2}}\,v$, and after performing the integration we then find that

$$g(u) = \frac{1}{\pi}\left(\frac{1}{1+u^2}\right).$$

Which is the distribution function for a Cauchy random variable.

Exercise 35 (a geometric example with the Cauchy distribution)

From the geometry of the problem we note that $\tan(\theta) = \frac{y}{1} = y$. Here θ is a uniform random variable with a domain of $\left(-\frac{\pi}{2}, +\frac{\pi}{2}\right)$ and we have a p.d.f. given by $g_\Theta(\theta) = \frac{1}{\pi}$. To transform this p.d.f from the random variable θ to the random variable y recall

$$g_Y(y) = g_\Theta(\theta) \left| \frac{\partial \theta}{\partial y} \right| .$$

From the form $y = \tan(\theta)$ we have $\frac{dy}{d\theta} = \sec^2(\theta)$, so $\frac{d\theta}{dy} = \frac{1}{\sec^2(\theta)}$. With this and using the identity $\sec^2(\theta) = 1 + \tan^2(\theta)$, we see that the p.d.f $g_Y(y)$ becomes

$$g_Y(y) = \frac{g_\Theta(\theta)}{1+\tan^2(\theta)} = \frac{1}{\pi}\left(\frac{1}{1+y^2}\right) ,$$

which is a Cauchy distribution.

Exercise 36 (the ratio $\frac{X/\alpha}{Y/\beta}$ has an F-distribution)

Let X be a χ^2 random variable with α degrees of freedom, and Y be a χ^2 random variable with β degrees of freedom, then we claim that the expression

$$\frac{X/\alpha}{Y/\beta} , \tag{39}$$

is a F distributed random variable with parameters α and β. To show this introduce the random variable $V = \frac{X/\alpha}{Y/\beta}$ and lets attempt to derive its distribution function. We have

$$\begin{aligned}
F_V(l) &= \Pr\{V < l\} = \Pr\{X < \frac{\alpha}{\beta} Y l\} \\
&= \int_{y=0}^{\infty} \int_{x=0}^{\frac{\alpha}{\beta} yl} p_Y(y) p_X(x) dx dy \\
&= \int_{y=0}^{\infty} \int_{x=0}^{\frac{\alpha}{\beta} yl} \frac{\left(\frac{1}{2}\right)^{\beta/2}}{\Gamma(\beta/2)} y^{\frac{\beta}{2}-1} e^{-\frac{1}{2}y} \frac{\left(\frac{1}{2}\right)^{\alpha/2}}{\Gamma(\alpha/2)} x^{\frac{\alpha}{2}-1} e^{-\frac{1}{2}x} dx dy \\
&= \left(\frac{1}{2}\right)^{\frac{\alpha+\beta}{2}} \frac{1}{\Gamma(\beta/2)\Gamma(\alpha/2)} \int_{y=0}^{\infty} y^{\frac{\beta}{2}-1} e^{-\frac{1}{2}y} \int_{x=0}^{\frac{\alpha}{\beta} yl} x^{\frac{\alpha}{2}-1} e^{-\frac{1}{2}x} dx dy .
\end{aligned}$$

Since we cannot evaluate the inner integral in closed form we will take the derivative of $F_V(l)$ with respect to l and derive the probability density function $f_V(l)$. We find

$$\begin{aligned}
f_V(l) &= F_V'(l) = \left(\frac{1}{2}\right)^{\frac{\alpha+\beta}{2}} \frac{1}{\Gamma(\alpha/2)\Gamma(\beta/2)} \int_{y=0}^{\infty} y^{\frac{\beta}{2}-1} e^{-\frac{1}{2}y} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}-1} y^{\frac{\alpha}{2}-1} l^{\frac{\alpha}{2}-1} e^{-\frac{1}{2}\frac{\alpha}{\beta} yl} \cdot \frac{\alpha}{\beta} y \cdot dy \\
&= \left(\frac{1}{2}\right)^{\frac{\alpha+\beta}{2}} \frac{l^{\frac{\alpha}{2}-1}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}} \int_{y=0}^{\infty} y^{\frac{\alpha}{2}+\frac{\beta}{2}-1} e^{-\frac{1}{2}\left(1+\frac{\alpha}{\beta} l\right) y} dy .
\end{aligned}$$

To evaluate this integral let $v = \frac{1}{2}\left(1 + \frac{\alpha}{\beta} l\right) y$, then $dv = \frac{1}{2}\left(1 + \frac{\alpha}{\beta} l\right) dy$, and we get

$$
\begin{aligned}
f_V(l) &= \left(\frac{1}{2}\right)^{\frac{\alpha+\beta}{2}} \frac{l^{\frac{\alpha}{2}-1}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \left(\frac{\alpha}{\beta}\right)^{\frac{\alpha}{2}} \frac{1}{\left(\frac{1}{2}\left(1+\frac{\alpha}{\beta} l\right)\right)^{\frac{\alpha+\beta}{2}}} \int_{v=0}^{\infty} v^{\frac{\alpha}{2}+\frac{\beta}{2}-1} e^{-v} dv \\
&= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \cdot \frac{l^{\alpha/2-1}}{(\beta+\alpha l)^{\frac{\alpha+\beta}{2}}},
\end{aligned}
$$

or an F distribution with α and β degrees of freedom.

Exercise 37 (transformation of F distributed random variables)

Part (a): Consider the transformation $y = \frac{1}{x}$, then

$$
\frac{dy}{dx} = -\frac{1}{x^2} \quad \text{so} \quad \frac{dx}{dy} = -x^2 .
$$

Thus

$$
\begin{aligned}
g_Y(y) &= g_X(x(y)) \left|\frac{dx}{dy}\right| = g_X(x(y)) \left|-x(y)^2\right| \\
&= \frac{1}{y^2} \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \frac{y^{-\alpha/2+1}}{(\beta+\alpha/y)^{\frac{\alpha+\beta}{2}}} \\
&= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \frac{y^{\beta/2-1}}{(\alpha+\beta y)^{\frac{\alpha+\beta}{2}}},
\end{aligned}
$$

which is an F distribution with parameters β and α as we were to show.

Part (b): Next consider the transformation $y = x^2$, then

$$
\frac{dy}{dx} = 2x \quad \text{so} \quad \frac{dx}{dy} = \frac{1}{2x} = \frac{1}{2\sqrt{y}} .
$$

So the p.d.f of y is given by

$$
\begin{aligned}
g_Y(y) &= g_X(x(y)) \left|\frac{dx}{dy}\right| \\
&= \frac{1}{2\sqrt{y}} g_X(x(y)) \\
&= \frac{1}{2\sqrt{y}} \frac{\Gamma((\alpha+1)/2)}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)} \left(1+\frac{y}{\alpha}\right)^{-\left(\frac{\alpha+1}{2}\right)} \\
&= \frac{\Gamma((\alpha+1)/2)\alpha^{\alpha/2}}{(\alpha\pi)^{1/2}\Gamma(\alpha/2)} \frac{y^{-1/2}}{(\alpha+y)^{\frac{\alpha+1}{2}}},
\end{aligned}
$$

which is an F distribution with parameters 1 and α degrees of freedom as we were to show.

Exercise 38 (transformation of beta distributed random variables)

Consider the variable Y defined in terms of X by

$$y = \frac{\alpha x}{\beta + \alpha x} \,. \tag{40}$$

Then we have the derivative of y with respect to x given by

$$\frac{dy}{dx} = \frac{\alpha}{\beta + \alpha x} - \frac{\alpha^2 x}{(\beta + \alpha x)^2} = \frac{\alpha\beta}{(\beta + \alpha x)^2} \,,$$

and

$$\frac{dx}{dy} = \frac{(\beta + \alpha x)^2}{\alpha\beta} \,.$$

From Equation 40 we have that x in terms of y is given by

$$x = \left(\frac{\beta}{\alpha}\right) \frac{y}{1 - y} \,. \tag{41}$$

Then since X is distributed as an F random variable with parameters α and β its density function looks like

$$g_X(x) = \frac{\Gamma(\frac{\alpha}{2} + \frac{\beta}{2})}{\Gamma(\alpha/2)\Gamma(\beta/2)} \alpha^{\alpha/2} \beta^{\beta/2} \frac{x^{\alpha/2-1}}{(\beta + \alpha x)^{(\alpha+\beta)/2}} \,. \tag{42}$$

When X is given by Equation 41 we see that the term in the denominator above looks like

$$\beta + \alpha x = \beta \left(\frac{1}{1 - y}\right) \,,$$

so

$$\frac{x^{\alpha/2-1}}{(\beta + \alpha x)^{(\alpha+\beta)/2}} = \frac{\beta^{-\beta/2-1}}{\alpha^{\alpha/2-1}} \frac{y^{\alpha/2-1}}{(1 - y)^{-\beta/2-1}} \,.$$

In addition, the derivative is given by

$$\left|\frac{dx}{dy}\right| = \frac{1}{\alpha\beta} \left(\beta + \beta \frac{y}{1 - y}\right)^2 = \frac{\beta}{\alpha} \left(\frac{1}{1 - y}\right)^2 \,.$$

Thus when we put everything together we get

$$\begin{aligned}
g_Y(y) &= g_X(x(y)) \left|\frac{dx}{dy}\right| \\
&= \frac{\Gamma(\alpha/2 + \beta/2)}{\Gamma(\alpha/2)\Gamma(\beta/2)} \alpha^{\alpha/2} \beta^{\beta/2} \frac{\beta^{-\beta/2-1}}{\alpha^{\alpha-1}} \frac{y^{\alpha/2-1}}{(1 - y)^{-\beta/2-1}} \frac{\beta}{\alpha} \frac{1}{(1 - y)^2} \\
&= \frac{\Gamma(\alpha/2 + \beta/2)}{\Gamma(\alpha/2)\Gamma(\beta/2)} y^{\frac{\alpha}{2}-1} (1 - y)^{\frac{\beta}{2}-1} \,,
\end{aligned}$$

which is a beta distribution with parameters $\frac{\alpha}{2}$ and $\frac{\beta}{2}$ as we were to show.

Exercise 39 (expectations of F distributed random variables)

To have $E(X^k)$ converge when X is given by a F distribution with parameters α and β given by Equation 42 requires that the integral of an expression like

$$\frac{x^k x^{\alpha/2-1}}{(\beta+\alpha x)^{\frac{\alpha+\beta}{2}}} \sim x^{k-\beta/2-1},$$

converge. This requires that as a function of x is has an exponent less than -1, which means that $k - \frac{\beta}{2} - 1 < -1$ or solving for β that

$$\beta > 2k. \tag{43}$$

Next we compute the expectations $E(X)$ and $E(X^2)$ using the functional form for the p.d.f. of an F distributed random variable with parameters α and β given by Equation 42. For $E(X)$ we have

$$\begin{aligned}
E(X) &= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \int_{x=0}^{\infty} \frac{x^{\alpha/2}}{(\beta+\alpha x)^{(\alpha+\beta)/2}} dx \\
&= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \left(\frac{\alpha^{-1-\alpha/2}\beta^{1-\beta/2}\Gamma(1+\alpha/2)\Gamma(-1+\beta/2)}{\Gamma(\alpha/2+\beta/2)} \right) \\
&= \left(\frac{\beta}{\alpha}\right) \frac{\Gamma(1+\alpha/2)}{\Gamma(\alpha/2)} \frac{\Gamma(-1+\beta/2)}{\Gamma(\beta/2)} \\
&= \left(\frac{\beta}{\alpha}\right) \frac{(\alpha/2)\,\Gamma(\alpha/2)}{\Gamma(\alpha/2)} \frac{\Gamma(-1+\beta/2)}{(\beta/2-1)\,\Gamma(\beta/2-1)} \\
&= \frac{\beta}{2} \frac{1}{(\beta/2-1)} = \frac{\beta}{\beta-2}.
\end{aligned}$$

Note that by Equation 43 to have the expectation integral converge requires that $\beta > 2$. For $E(X^2)$ we have

$$\begin{aligned}
E(X^2) &= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \int_{x=0}^{\infty} \frac{x^{\alpha/2+1}}{(\beta+\alpha x)^{(\alpha+\beta)/2}} dx \\
&= \frac{\Gamma(\frac{\alpha+\beta}{2})\alpha^{\alpha/2}\beta^{\beta/2}}{\Gamma(\alpha/2)\Gamma(\beta/2)} \left(\frac{\alpha^{-2-\alpha/2}\beta^{2-\beta/2}\Gamma(2+\alpha/2)\Gamma(-2+\beta/2)}{\Gamma((\alpha+\beta)/2)} \right) \\
&= \frac{\beta^2}{\alpha^2} \frac{(1+\alpha/2)(\alpha/2)\Gamma(\alpha/2)}{\Gamma(\alpha/2)} \cdot \frac{\Gamma(-2+\beta/2)}{(-1+\beta/2)(-2+\beta/2)\Gamma(-2+\beta/2)} \\
&= \frac{\beta^2(\alpha+2)}{\alpha(\beta-2)(\beta-4)}.
\end{aligned}$$

Again note that by Equation 43 to have the squared expectation of X integral converge requires that $\beta > 4$. To compute the variance we use the standard formula involving $E(X^2)$ and $E(X)$ of

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{2\beta^2(\alpha+\beta-2)}{\alpha(\beta-4)(\beta-2)^2},$$

when we use the above expressions. The needed integrations for this problem can be found in the Mathematica notebook `chap_4_prob_39.nb`.

Chapter 5 (Some special multivariate distributions)

Notes on the text

Notes on the Dirichlet distribution

Given the result that for a Dirichlet distribution the expectation of powers of the random variables is simple to compute using

$$E(X_1^{r_1}\cdots X_k^{r_k}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \cdot \frac{\prod_{i=1}^k \Gamma(\alpha_i + r_i)}{\Gamma[\sum_{i=1}^k (\alpha_i + r_i)]}, \tag{44}$$

we can compute the simplest statistics for components of the Dirichlet distribution. To simply the notation in the following we define $\alpha_0 = \sum_{i=1}^k \alpha_k$. Then we can compute some simple statistics of the components of X when X is distributed with a Dirichlet distribution. For example to evaluate $E(X_i)$ we have $r_i = 1$ and $r_j = 0$ for all $j \neq i$. Then we find that

$$E(X_i) = \frac{\Gamma(\alpha_0)\Gamma(\alpha_i+1)\prod_{j=1;j\neq i}^k \Gamma(\alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)\Gamma(\alpha_0+1)} = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_0+1)} \cdot \frac{\Gamma(\alpha_i+1)}{\Gamma(\alpha_i)}.$$

Since $\Gamma(x+1) = x\Gamma(x)$ the above becomes

$$E(X_i) = \frac{\alpha_i}{\alpha_0}, \tag{45}$$

which is equation 6 in this section. To compute $\mathrm{Var}(X_i)$ we need to compute $E(X_i^2)$. Using the same technique as to compute $E(X_i)$ we have

$$\begin{aligned} E(X_i^2) &= \frac{\Gamma(\alpha_0)}{\prod_{j=1}^k \Gamma(\alpha_j)} \cdot \frac{\prod_{j=1;j\neq i}^k \Gamma(\alpha_j)\cdot\Gamma(\alpha_i+2)}{\Gamma(\alpha_0+2)} \\ &= \frac{\Gamma(\alpha_0)}{(\alpha_0+1)\alpha_0\Gamma(\alpha_0)} \cdot \frac{\Gamma(\alpha_i+2)}{\Gamma(\alpha_i)} = \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)}. \end{aligned} \tag{46}$$

Using this we can compute $\mathrm{Var}(X_i)$ as

$$\mathrm{Var}(X_i) = E(X_i^2) - E(X_i)^2 = \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} - \frac{\alpha_i^2}{\alpha_0^2} = \frac{\alpha_i(\alpha_0-\alpha_i)}{\alpha_0^2(\alpha_0+1)}, \tag{47}$$

which is the result quoted in the book. The final result we will derive is $\mathrm{Cov}(X_i, X_j)$ to compute this we need $E(X_iX_j)$. We find

$$\begin{aligned} E(X_iX_j) &= \frac{\Gamma(\alpha_0)\prod_{l=1;l\neq i,j}^k \Gamma(\alpha_l)\Gamma(\alpha_i+1)\Gamma(\alpha_j+1)}{\prod_{l=1}^k \Gamma(\alpha_l)\Gamma(\alpha_0+2)} \\ &= \frac{1}{\alpha_0(\alpha_0+1)} \cdot \frac{\Gamma(\alpha_i+1)}{\Gamma(\alpha_i)} \cdot \frac{\Gamma(\alpha_j+1)}{\Gamma(\alpha_j)} \\ &= \frac{\alpha_i\alpha_j}{\alpha_0(\alpha_0+1)}. \end{aligned} \tag{48}$$

Thus we find

$$
\begin{aligned}
\text{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) = \frac{\alpha_i \alpha_j}{\alpha_0(\alpha_0 + 1)} - \frac{\alpha_i \alpha_j}{\alpha_0^2} \\
&= -\frac{\alpha_i \alpha_j}{\alpha_0^2(\alpha_0 + 1)}\,, \qquad (49)
\end{aligned}
$$

the same as claimed in the book.

Notes on the multivariate t distribution

Recall that if we know the p.d.f of the vector $(\mathbf{y}, z)$ but want the p.d.f of the vector $(\mathbf{x}, z)$ it is given by

$$
g_{\mathbf{X},Z}(\mathbf{x}, z) = g_{\mathbf{Y},Z}(\mathbf{y}, z) \left| \frac{\partial(\mathbf{y}, z)}{\partial(\mathbf{x}, z)} \right| .
$$

To use this we need to compute

$$
\frac{\partial(\mathbf{y}, z)}{\partial(\mathbf{x}, z)} = \begin{bmatrix}
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_k} & \frac{\partial y_1}{\partial z} \\
\frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_k} & \frac{\partial y_2}{\partial z} \\
\vdots & & & & \vdots \\
\frac{\partial y_k}{\partial x_1} & \frac{\partial y_k}{\partial x_2} & \cdots & \frac{\partial y_k}{\partial x_k} & \frac{\partial y_k}{\partial z} \\
\frac{\partial z}{\partial x_1} & \frac{\partial z}{\partial x_2} & \cdots & \frac{\partial z}{\partial x_k} & \frac{\partial z}{\partial z}
\end{bmatrix} .
$$

Since $Y_i = \left(\frac{Z}{n}\right)^{1/2}(X_i - \mu)$ we see that

$$
\begin{aligned}
\frac{\partial Y_i}{\partial x_j} &= \begin{cases} 0 & i \neq j \\ \left(\frac{z}{n}\right)^{1/2} & i = j \end{cases} \\
\frac{\partial Y_i}{\partial z} &= \frac{1}{2}\frac{1}{\sqrt{n}} z^{-1/2}(X_i - \mu_i) \quad \forall \quad i \\
\frac{\partial z}{\partial x_i} &= 0 \quad \forall \quad i \\
\frac{\partial z}{\partial z} &= 1\,.
\end{aligned}
$$

Putting these expressions into $\frac{\partial(\mathbf{y},z)}{\partial(\mathbf{x},z)}$ we have

$$
\frac{\partial(\mathbf{y}, z)}{\partial(\mathbf{x}, z)} = \begin{bmatrix}
\left(\frac{z}{n}\right)^{1/2} & 0 & \cdots & 0 & \frac{1}{2}\frac{1}{\sqrt{n}} z^{-1/2}(X_1 - \mu_1) \\
0 & \left(\frac{z}{n}\right)^{1/2} & \cdots & 0 & \frac{1}{2}\frac{1}{\sqrt{n}} z^{-1/2}(X_2 - \mu_2) \\
\vdots & & & & \vdots \\
0 & 0 & \cdots & \left(\frac{z}{n}\right)^{1/2} & \frac{1}{2}\frac{1}{\sqrt{n}} z^{-1/2}(X_k - \mu_k) \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix} .
$$

From this expression we see that the determinant of this expression is given by $\left(\frac{z}{n}\right)^{k/2}$, as claimed in the book. Thus when we put in the known p.d.f for $g_{\mathbf{Y},Z}(\mathbf{y}, z)$ we have

$$
g_{\mathbf{X},Z}(\mathbf{x}, z) = g_{\mathbf{Y},Z}(\mathbf{y}, z) \left(\frac{z}{n}\right)^{k/2}
$$

$$
\begin{aligned}
&= (2\pi)^{-k/2}|T|^{1/2}\exp\left\{-\frac{1}{2}\left(\frac{z}{n}\right)(x-\mu)'T(x-\mu)\right\} \\
&\times \left[2^{n/2}\Gamma\left(\frac{n}{2}\right)\right]^{-1} z^{(n/2)-1}e^{-z/2}\left(\frac{z}{n}\right)^{k/2} \\
&= \frac{|T|^{1/2}}{(2\pi n)^{k/2}2^{n/2}\Gamma\left(n/2\right)}\exp\left\{-\frac{1}{2}\left(\frac{z}{n}\right)(x-\mu)'T(x-\mu)\right\}. \qquad (50)
\end{aligned}
$$

or the books equation 6. Since we ultimately want the p.d.f of X by itself we need to integrate Z out of $g_{\mathbf{X},Z}(\mathbf{x},z)$. To do this we will use the following integration identity

$$
\int_0^\infty z^{(n+k-2)/2}e^{-Qz}dz = \Gamma\left(\frac{n+k}{2}\right)Q^{-(n+k)/2}. \qquad (51)
$$

To prove this equation let $\xi = Qz$ so that that $z = \frac{\xi}{Q}$ and $d\xi = Qdz$ and the left-hand-side of Equation 51 to get

$$
\int_0^\infty \frac{\xi^{(n+k-2)/2}}{Q^{(n+k-2)/2}}e^{-\xi}\left(\frac{d\xi}{Q}\right) = \frac{1}{Q^{(n+k)/2}}\int_0^\infty \xi^{(n+k)/2-1}e^{-\xi}d\xi = \Gamma\left(\frac{n+k}{2}\right)Q^{-(n+k)/2},
$$

which is the desired expression. Thus using this integral and the books notation for the constant c' we can compute $g_{\mathbf{X}}(\mathbf{x}|n,\mu,T)$ as

$$
\begin{aligned}
g_{\mathbf{X}}(\mathbf{x}|n,\mu,T) &= \int_0^\infty g_{\mathbf{X},Z}(\mathbf{x},z)dz \\
&= c'\int_0^\infty z^{(n+k-2)/2}e^{-\frac{1}{2}(1+\frac{1}{n}(x-\mu)'T(x-\mu))z}dz \\
&= c'\Gamma\left(\frac{n+k}{2}\right)\left[1+\frac{1}{n}(x-\mu)'T(x-\mu)\right]^{-\left(\frac{n+k}{2}\right)},
\end{aligned}
$$

the same expression as in the book. In summary then, the p.d.f of a k-dimensional multivariate t-distribution with n degrees of freedom, a location vector μ, and a precision matrix T is given by

$$
g_{\mathbf{X}}(\mathbf{x}|n,\mu,T) = \frac{\Gamma\left(\frac{n+k}{2}\right)|T|^{1/2}}{\Gamma\left(\frac{n}{2}\right)(n\pi)^{k/2}}\left[1+\frac{1}{n}(x-\mu)'T(x-\mu)\right]^{-\left(\frac{n+k}{2}\right)}. \qquad (52)
$$

Notes on the bilateral bivariate Pareto distribution

Since the computed marginal distributions for X_1 and X_2 derived in Exercise 23 and expressed by Equations 98 and 99 are univariate Pareto distribution we can use the expectation for a univariate Pareto distribution given by Equation 34 to derive the expectation of the marginals of the bivariate Pareto. As such, since $r_2 - X_1$ is a univariate Pareto from Equation 34 we have that

$$
E(r_2 - X_1) = \frac{\alpha(r_2 - r_1)}{\alpha - 1},
$$

or since r_1 is a constant in the expectation this becomes

$$
r_2 - E(X_1) = \frac{\alpha(r_2 - r_1)}{\alpha - 1},
$$

so solving for $E(X_1)$ we find

$$E(X_1) = r_2 - \frac{\alpha(r_2 - r_1)}{\alpha - 1} = \frac{\alpha r_1 - r_2}{\alpha - 1}. \tag{53}$$

Using the expression $E(X^2)$ for a univariate Pareto distribution given by Equation 35 on the variable $r_2 - X_1$ means in this case that

$$E((r_2 - X_1)^2) = \frac{\alpha(r_2 - r_1)^2}{\alpha - 2}.$$

Expanding the quadratic expression on the left-hand-side of the above gives

$$r_2^2 - 2r_2 E(X_1) + E(X_1^2) = \frac{\alpha(r_2 - r_1)^2}{\alpha - 2}.$$

Putting in the known value of $E(X_1)$ given by Equation 53 above gives the following expression for $E(X_1^2)$.

$$r_2^2 - 2r_2 \left(\frac{\alpha r_1 - r_2}{\alpha - 1} \right) + E(X_1^2) = \frac{\alpha(r_2 - r_1)^2}{\alpha - 2}.$$

When we solve this expression for $E(X_1^2)$ we get

$$E(X_1^2) = \frac{\alpha(\alpha - 1)r_1^2 - 2\alpha r_1 r_2 + 2r_2^2}{(\alpha - 1)(\alpha - 2)}. \tag{54}$$

From which we can get that $\text{Var}(X_1)$ is given by

$$\text{Var}(X_1) = E(X_1^2) - E(X_1)^2 = \frac{\alpha(r_2 - r_1)^2}{(\alpha - 1)^2(\alpha - 2)}. \tag{55}$$

Next we will perform the same manipulations as above but for the marginal distribution of X_2. Since from Exercise 23 and using Equation 34 we have that

$$E(X_2 - r_1) = \frac{\alpha(r_2 - r_1)}{\alpha - 1}.$$

or since r_1 is a constant we can solve for $E(X_2)$ to get

$$E(X_2) = r_1 + \frac{\alpha(r_2 - r_1)}{\alpha - 1} = \frac{\alpha r_2 - r_1}{\alpha - 1}. \tag{56}$$

Using $E(X^2)$ for a univariate Pareto distribution given by Equation 35 means in this case that

$$E((X_2 - r_1)^2) = E(X_2^2) - 2r_1 E(X_2) + r_1^2 = \frac{\alpha(r_2 - r_1)^2}{\alpha - 2}.$$

Putting in the known value of $E(X_2)$ given by Equation 56 gives the following expression we must solve for $E(X_2^2)$.

$$E(X_2^2) - 2r_1 \left(\frac{\alpha r_2 - r_1}{\alpha - 1} \right) + r_1^2 = \frac{\alpha(r_2 - r_1)^2}{\alpha - 2}.$$

We can solve this for $E(X_2^2)$ to get

$$E(X_2^2) = \frac{\alpha(\alpha-1)r_1^2 - 2\alpha r_1 r_2 + 2r_1^2}{(\alpha-1)(\alpha-2)}\,, \tag{57}$$

and then use this value to get $\text{Var}(X_2)$, where we find

$$\text{Var}(X_2) = E(X_2^2) - E(X_2)^2 = \frac{\alpha(r_2-r_1)^2}{(\alpha-1)^2(\alpha-2)}\,. \tag{58}$$

Some of the algebra for these problems is worked in the Mathematica notebook `bilateral_Pareto_Derivations.nb`.

Exercise Solutions

Exercise 1 (independent Poisson random variables that sum to n)

For this problem we are asked to evaluate $P(X|\sum_{i=1}^k x_i = n)$. To evaluate this recall the definition of conditional probability

$$P(X|\sum_{i=1}^k x_i = n) = \frac{P(X, \sum_{i=1}^k x_i = n)}{P(\sum_{i=1}^k x_i = n)}\,.$$

Since each component of the vector X is an independent Poisson random variable we can compute the expression $P(X, \sum_{i=1}^k x_i = n)$ in terms of products of the densities of the components x_i as

$$\left(\prod_{i=1}^{k-1}\left(\frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}\right)\right)\left(\frac{e^{-\lambda_k}\lambda_k^{n-\sum_{j=1}^{k-1}x_j}}{(n-\sum_{j=1}^{k-1}x_j)!}\right)\,.$$

If we introduce the non-random variable x_k defined in terms of the earlier variables x_i for $1 \le i \le k-1$ as $x_k \equiv n - \sum_{j=1}^{k-1} x_j$, the above simplifies to

$$P(X, \sum_{i=1}^k x_i = n) = \prod_{i=1}^k \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}\,.$$

Next we need to evaluate $P(\sum_{i=1}^k x_i = n)$. This is greatly simplified if we recall that the *sum* of independent individual Poisson random variables with parameters λ_i is another Poisson random variable with parameter $\sum_{i=1}^k \lambda_i$. Thus

$$P(\sum_{i=1}^k x_i = n) = \frac{e^{-\sum_{i=1}^k \lambda_i}(\sum_{i=1}^k \lambda_i)^n}{n!}\,.$$

Using these two results we find

$$\begin{aligned} P(X|\sum_{i=1}^k x_i = n) &= \left(\prod_{i=1}^k \frac{e^{-\lambda_i}\lambda_i^{x_i}}{x_i!}\right)\cdot\frac{n!}{e^{-\sum_{i=1}^k \lambda_i}(\sum_{i=1}^k \lambda_i)^n} \\ &= \frac{n!}{x_1!x_2!\cdots x_k!}\left(\frac{\prod_{i=1}^k \lambda_i^{x_i}}{(\sum_{i=1}^k \lambda_i)^n}\right)\,. \end{aligned}$$

Since $n = \sum_{i=1}^k x_i$ we can write

$$(\sum_{j=1}^k \lambda_j)^n = (\sum_{j=1}^k \lambda_j)^{\sum_{i=1}^k x_i} = \prod_{i=1}^k (\sum_{j=1}^k \lambda_j)^{x_i} .$$

Thus using this $P(X|\sum_{i=1}^k x_i = n)$ becomes

$$\begin{aligned} P(X|\sum_{i=1}^k x_i = n) &= \frac{n!}{x_1!x_2!\cdots x_k!}\left(\frac{\prod_{i=1}^k \lambda_i^{x_i}}{\prod_{i=1}^k (\sum_{j=1}^k \lambda_j)^{x_i}}\right) \\ &= \frac{n!}{x_1!x_2!\cdots x_k!}\prod_{i=1}^k \left(\frac{\lambda_i}{(\sum_{j=1}^k \lambda_j)}\right)^{x_i} , \end{aligned}$$

which is the P.D.F. of a multinomial distribution with probability p_i given by

$$p_i = \frac{\lambda_i}{(\sum_{j=1}^k \lambda_j)} ,$$

as we were to show.

Exercise 2 (the characteristic function for a multinomial distribution)

The characteristic of an n-dimensional random variable $X = (X_1, X_2, \cdots, X_k)'$ is a complex valued function of defined at each point $t = (t_1, t_2, \cdots, t_k)'$ given by

$$\zeta(t) = E\left(e^{it'X}\right) . \tag{59}$$

In this case this becomes

$$\begin{aligned} \zeta(t) &= \sum_{x_1=0}^n \sum_{x_2=0}^n \cdots \sum_{x_k=0}^n \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} e^{it'X} \\ &= \sum_{x_1=0}^n \sum_{x_2=0}^n \cdots \sum_{x_k=0}^n \frac{n!}{x_1!x_2!\cdots x_k!} (p_1 e^{it_1})^{x_1} (p_2 e^{it_2})^{x_2} \cdots (e^{it_k} p_k)^{x_k} . \end{aligned}$$

Using the multinomial formula this becomes

$$\zeta(t) = \left(\sum_{j=1}^k p_j e^{it_j}\right)^n , \tag{60}$$

as we were to show.

To compute $E(X)$ lets consider $E(X_i)$ the componentwise expectation for $i = 1, 2, \cdots, p$ then

$$E(X_i) = \sum_{(x_1, \cdots x_i, \cdots, x_k)} x_i \left(\begin{array}{c} n \\ x_1\,, x_2\,, \cdots\,, x_k \end{array}\right) p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} .$$

We can simplify in this case the coefficient in the above expression as follows

$$\begin{aligned} x_i \binom{n}{x_1\,, x_2 \cdots x_k} &= x_i \frac{n!}{x_1! \cdot x_i! \cdots x_k!} \\ &= \frac{n(n-1)!}{x_1! \cdots (x_i - 1)! \cdots x_k!} \\ &= n \frac{(n-1)!}{x_1! \cdots (x_i - 1)! \cdots x_k!} \,. \end{aligned}$$

Writing $p_i^{x_i} = p_i p_i^{x_i - 1}$ we can express $E(X_i)$ as

$$E(X_i) = np_i \sum_X \binom{n-1}{x_1\,, \cdots\,, x_i - 1\,, \cdots\,, x_k} p_1^{x_1} \cdots p_2^{x_i - 1} \cdots p_k^{x_k} \,.$$

Where the sum is over the vectors $X = (x_1, \cdots x_i - 1, \cdots, x_k)'$ such that

$$x_1 + x_2 + \cdots + x_i - 1 + \cdots + x_k = n - 1 \,.$$

To simplify this, simply rename the $x_i - 1$ variable something like $\tilde{x}_i$ and we now have sum

$$E(X_i) = np_i \sum_X \binom{n-1}{x_1\,, \cdots\,, \tilde{x}_i\,, \cdots\,, x_k} p_1^{x_1} \cdots p_2^{\tilde{x}_i} \cdots p_k^{x_k} \,.$$

The sum above evaluates to one and we get

$$E(X_i) = np_i \,. \tag{61}$$

In vector form this, equation is $E(X) = np$. We can also argue the correctness of this expression by defining an event E_i to have occurred when we increment our ith counting random variable X_i from X_i to $X_i + 1$. Then the event E_i happens with probability p_i and does not happen with probability $1 - p_i$. When dealing with the event E_i by itself, the variable X_i, representing the number of events E_i that happen in n trials is a *binomial* random variable, and has the known expectation expression given by Equation 61.

In computing the variance recall that $\text{Var}(X_i) = E(X_i^2) - E(X_i)^2 = E(X_i^2) - (np_i)^2$, thus to evaluate $\text{Var}(X_i)$ we need to be able to evaluate $E(X_i^2)$. This later expectation is given by

$$E(X_i^2) = \sum_{|x|=n} x_i^2 \binom{n}{x_1\,, x_2\,, \cdots x_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k}$$

In this sum we have used the notation $|x| = n$ to mean the vector of elements of x must have components that sum to n or

$$x_1 + x_2 + \cdots + x_k = n \,. \tag{62}$$

Note that in the above we can separate this sum out into two parts. The first where $x_i = 0$ and the second where $x_i \neq 0$ as

$$\begin{aligned} E(X_i^2) &= \sum_{|x|=n;\, x_i=0} x_i^2 \binom{n}{x_1\,, x_2\,, \cdots x_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k} \\ &+ \sum_{|x|=n;\, x_i \neq 0} x_i^2 \binom{n}{x_1\,, x_2\,, \cdots x_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k} \,, \end{aligned}$$

But this first sum is zero by definition leaving just the second sum. In this second sum consider the leading coefficient. Since $x_i \neq 0$ we have

$$\begin{aligned} x_i^2 \binom{n}{x_1\,, x_2\,, \cdots x_i\,, \cdots x_k} &= x_i^2 \frac{n!}{x_1! x_2! \cdots x_i! \cdots x_k!} \\ &= n x_i \frac{(n-1)!}{x_1! x_2! \cdots (x_i-1)! \cdots x_k!}\,, \end{aligned}$$

so our expectation expression becomes

$$E(X_i^2) = n \sum_{|x|=n;\, x_i \neq 0} x_i \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k}\,.$$

Lets write the variable x_i as $x_i - 1 + 1$ and then split the above single sum into two sums as

$$\begin{aligned} E(X_i^2) &= n \sum_{|x|=n; x_i \neq 0} (x_i - 1) \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k} \quad (63) \\ &+ n \sum_{|x|=n; x_i \neq 0} \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k}\,. \quad (64) \end{aligned}$$

For this first sum we can use the same trick as before in that we can break this sum up into two pieces, in this case sums when $x_i = 1$ and sums when $x_i \neq 1$. Since all the terms have a coefficient $x_i - 1$ we see that when we perform this decomposition the sum over the points where $x_i = 1$ has each term vanish and we can write $E(X_i^2)$ as

$$\begin{aligned} E(X_i^2) &= n \sum_{|x|=n; x_i \neq \{0,1\}} (x_i - 1) \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k} \\ &+ n \sum_{|x|=n; x_i \neq 0} \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k}\,. \end{aligned}$$

The coefficient of the expression in the first sum (since $x_i \neq 1$) can be written as

$$\begin{aligned} (x_i - 1) \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} &= (x_i - 1) \frac{(n-1)!}{x_1! x_2! \cdots (x_i-1)! \cdots x_k!} \\ &= (n-1) \frac{(n-2)!}{x_1! x_2! \cdots (x_i-2)! \cdots x_k!}\,, \end{aligned}$$

and we get for $E(X_i^2)$ the following

$$\begin{aligned} E(X_i^2) &= n(n-1) \sum_{|x|=n; x_i \neq \{0,1\}} \binom{n-2}{x_1\,, x_2\,, \cdots (x_i-2)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k} \quad (65) \\ &+ n \sum_{|x|=n; x_i \neq 0} \binom{n-1}{x_1\,, x_2\,, \cdots (x_i-1)\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{x_i} \cdots p_k^{x_k}\,. \quad (66) \end{aligned}$$

We'll now show how to evaluate these two remaining sums. For the second sum given by 66 above recognize that the constraint that $|x| = n$ and $x_i \neq 0$ means that none of the other values of x_j ($j \neq i$) can equal n, since if they did the sum constraint would require $x_i = 0$, which is forbidden. Thus the range of the other variables become $0 \leq x_j \leq n-1$, for $j \neq i$. Next in writing constraint $|x| = n$ from Equation 62 requires

$$x_1 + x_2 + \cdots + (x_i - 1) + \cdots + x_k = n - 1\,.$$

If we introduce a variable $\tilde{x}_i = x_i - 1$, this constraint becomes

$$x_1 + x_2 + \cdots + \tilde{x}_i + \cdots + x_k = n - 1\,.$$

where $\tilde{x}_i$ (since $x_i \neq 0$) now takes the range $0 \leq \tilde{x}_i \leq n-1$ as all the others. Thus we end with the sum

$$n \sum_{|\tilde{x}|=n-1} \binom{n-1}{x_1\,, x_2\,, \cdots \tilde{x}_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{\tilde{x}_i+1} \cdots p_k^{x_k}\,,$$

or

$$np_i \sum_{|\tilde{x}|=n-1} \binom{n-1}{x_1\,, x_2\,, \cdots \tilde{x}_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{\tilde{x}_i} \cdots p_k^{x_k} = np_i\,,$$

Since the final remaining sum is evaluated as $(p_1 + p_2 + \cdots + p_k)^{n-1} = 1$ by the multinomial theorem. We can do this same trick with the first term given by 65 above. In this case the same logic show that since $x_i \neq 0$ we can introduce $\tilde{x}_i = x_i - 1$ and obtain the constraint

$$x_1 + x_2 + \cdots + \tilde{x}_i + \cdots + x_k = n - 1\,.$$

where all variables are restricted to the domain $[0, n-1]$. Since $x_i \neq 1$ also this means that $\tilde{x}_i \neq 0$ and thus $x_j \neq n-1$ so the above constraint requires that

$$x_1 + x_2 + \cdots + (\tilde{x}_i - 1) + \cdots + x_k = n - 2\,.$$

or

$$x_1 + x_2 + \cdots + \hat{x}_i + \cdots + x_k = n - 2\,.$$

where $\hat{x}_i = \tilde{x}_i - 1$ and *all* variables above constraint are now in the domain $[0, n-2]$. Thus our first sum becomes

$$n(n-1) \sum_{|\hat{x}|=n-2} \binom{n-2}{x_1\,, x_2\,, \cdots \hat{x}_i\,, \cdots x_k} p_1^{x_1} p_2^{x_2} \cdots p_i^{\hat{x}_i+2} \cdots p_k^{x_k} = n(n-1)p_i^2\,,$$

since after we factor our p_i^2 the sum above simplifies to $(p_1 + p_2 + \cdots + p_k)^{n-2} = 1^{n-2} = 1$ by the multinomial theorem. Combining these two expressions we finally arrive at

$$\text{Var}(X_i) = E(X_i^2) - E(X_i)^2 = n(n-1)p_i^2 + np_i - n^2 p_i^2 = np_i(1-p_i)\,, \tag{67}$$

as we were to show.

As the final part of this problem we will compute $\text{Cov}(X_i, X_j)$. From its definition we can show that $\text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$ so that given what we have already computed we need to now compute $E(X_i X_j)$ to evaluate this. This second expectation becomes

$$E(X_i X_j) \quad = \quad \sum_{|x|=n} x_i x_j \binom{n}{x_1\,, x_2\,, \cdots\,, x_i\,, \cdots\,, x_j\,, \cdots x_k} p_1^{x_1} \cdots p_i^{x_i} \cdots p_j^{x_j} \cdots p_k^{x_k}\,.$$

In this last sum we can restrict the values of x_i and x_j such that they are greater than or equal to one. We can write the multinomial coefficient in the above as

$$x_i x_j \binom{n}{x_1\,, x_2\,, \cdots\,, x_i\,, \cdots\,, x_j\,, \cdots x_k} = n(n{-}1)\binom{n-2}{x_1\,, x_2\,, \cdots\,, (x_i-1)\,, \cdots\,, (x_j-1)\,, \cdots x_k}\,.$$

As we have done before by introducing the variables $\tilde{x}_i = x_i - 1$ and $\tilde{x}_j = x_j - 1$ we find $E(X_iX_j)$ is given by

$$\begin{aligned} E(X_iX_j) &= n(n-1)\sum_{|\tilde{x}|=n-2}\binom{n-2}{x_1\,, x_2\,, \cdots\,, \tilde{x}_i\,, \cdots\,, \tilde{x}_j\,, \cdots x_k} p_1^{x_1}\cdots p_i^{\tilde{x}_i+1}\cdots p_j^{\tilde{x}_j+1}\cdots p_k^{x_k} \\ &= = n(n-1)p_ip_j\,. \end{aligned}$$

Using this expectation we can derive the value of $\text{Cov}(X_i, X_j)$ as

$$\text{Cov}(X_i, X_j) = n(n-1)p_ip_j - n^2p_ip_j = -np_ip_j\,, \tag{68}$$

as we were to show. Note that since from the expression for the covariance derived above is valid for $i \neq j$ if we sum across the ith row we have that

$$\sum_{j=1;j\neq i}^{n} \text{Cov}(X_i, X_j) = -\sum_{j=1;j\neq i}^{n} np_ip_j = -np_i(1-p_i)\,.$$

when we add this to Equation 67 we see that the row sum of the ith row for each i is zero. Thus since a non-zero combination of the columns of this matrix sums to the zero vector I would claim that the covariance matrix is *singular*.

Exercise 3 (adding independent multinomial variables with the same p vector)

Recall that the sum of independent random variables has a characteristic function that is the *product* of the characteristic functions for the individual random variables we have that

$$\zeta_{X_1+\cdots+X_r}(t) = \prod_{i=1}^{r}\zeta_{X_i}(t)\,,$$

where in this case we have $\zeta_{X_i}(t)$ given by Equation 60, with $n = n_i$. Using this expression we have

$$\zeta_{X_1+\cdots+X_r}(t) = \prod_{i=1}^{r}\left(\sum_{j=1}^{k} p_j e^{it_j}\right)^{n_i} = \left(\sum_{j=1}^{k} p_j e^{it_j}\right)^{\sum_{i=1}^{r} n_i}\,,$$

which is the characteristic function of a multinomial distribution with parameters $\sum_{i=1}^{r} n_i$ and the same probability vector $\mathbf{p}$ as all of the component multinomial random variables X_i.

Exercise 4 (the normalization of the Dirichlet distribution)

We will begin by attempting to transform the p.d.f of $\mathbf{X} \in R^{k-1}$ into one for $\mathbf{Y} \in R^{k-1}$. To do this we will use the general transformation between p.d.f's given by

$$g_{\mathbf{Y}}(\mathbf{y}) = g_{\mathbf{X}}(\mathbf{x}(\mathbf{y})) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| .$$

Note that for the given "direct" transformation between x_i and y_i we can derive the inverse transformation as

$$\begin{aligned}
x_1 &= y_1 \\
x_2 &= (1-x_1)y_2 = (1-y_1)y_2 \\
x_3 &= (1-x_1-x_2)y_3 = (1-y_1-(1-y_1)y_2)y_3 = (1-y_1)(1-y_2)y_3 = y_3\prod_{j=1}^{2}(1-y_j) \\
x_4 &= (1-y_1)(1-y_2)(1-y_3)y_4 = y_4\prod_{j=1}^{3}(1-y_j) \\
&\vdots \\
x_{k-1} &= (1-y_1)(1-y_2)(1-y_3)\cdots(1-y_{k-2})y_{k-1} = y_{k-1}\prod_{j=1}^{k-2}(1-y_j) .
\end{aligned}$$

Thus we find $\frac{\partial \mathbf{x}}{\partial \mathbf{y}}$ a $(k-1)\times(k-1)$ lower diagonal matrix given by

$$\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
-y_2 & 1-y_1 & 0 & \cdots & 0 \\
-(1-y_2)y_3 & -(1-y_1)y_3 & (1-y_1)(1-y_2) & \cdots & 0 \\
\vdots & & & & \vdots \\
-\left(\prod_{j=2}^{k-2}(1-y_j)\right)y_{k-1} & -\left(\prod_{j=1;j\neq 2}^{k-2}(1-y_j)\right)y_{k-1} & -\left(\prod_{j=1;j\neq 3}^{k-2}(1-y_j)\right)y_{k-1} & \cdots & \prod_{j=1}^{k-2}(1-y_j)
\end{bmatrix} .$$

The determinant of this matrix is the product of the elements on the diagonal and we find

$$\begin{aligned}
\left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| &= 1\cdot(1-y_1)\cdot(1-y_1)(1-y_2)\cdot(1-y_1)(1-y_2)(1-y_3)\cdots\prod_{j=1}^{k-2}(1-y_j) \\
&= (1-y_1)^{k-2}(1-y_2)^{k-3}(1-y_3)^{k-4}\cdots(1-y_{k-3})^2(1-y_{k-2})^1 \\
&= \prod_{j=1}^{k-1}(1-y_j)^{k-j-1} . \qquad (69)
\end{aligned}$$

Thus with this change of variables the integrand of $\mathcal{I} = \prod_{i=1}^{k} x_i^{\alpha_i - 1}$ our integral becomes

$$\begin{aligned}
\mathcal{I} &= y_1^{\alpha_1-1}\cdot(1-y_1)^{\alpha_2-1}y_2^{\alpha_2-1}\cdot(1-y_1)^{\alpha_3-1}(1-y_2)^{\alpha_3-1}y_3^{\alpha_3-1}\cdots \\
&\times \left(y_{k-1}\prod_{j=1}^{k-2}(1-y_j)\right)^{\alpha_{k-1}-1} x_k^{\alpha_k-1}\left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| d\mathbf{y} , \qquad (70)
\end{aligned}$$

Note that in the above expression we can write the variable x_k in terms of the variables $\mathbf{y}$ as

$$
\begin{aligned}
x_k &= 1 - x_1 - x_2 - x_3 - \cdots - x_{k-2} - x_{k-1} \\
&= 1 - y_1 - (1-y_1)y_2 - (1-y_1)(1-y_2)y_3 - \cdots \\
&- (1-y_1)(1-y_2)\cdots(1-y_{k-3})y_{k-2} - (1-y_1)(1-y_2)\cdots(1-y_{k-2})y_{k-1} \\
&= \prod_{j=1}^{k-1}(1-y_j)\,. \qquad (71)
\end{aligned}
$$

Thus using Equations 69 and 71 in Equation 70 the integrand $\mathcal{I}$ above becomes

$$
\begin{aligned}
\mathcal{I} &= y_1^{\alpha_1-1}\cdot(1-y_1)^{\alpha_2-1}y_2^{\alpha_2-1}\cdot(1-y_1)^{\alpha_3-1}(1-y_2)^{\alpha_3-1}y_3^{\alpha_3-1}\cdots\left(y_{k-1}\prod_{j=1}^{k-2}(1-y_j)\right)^{\alpha_{k-1}-1} \\
&\times \left(\prod_{j=1}^{k-1}(1-y_j)\right)^{\alpha_k-1}\times\prod_{j=1}^{k-1}(1-y_j)^{k-j-1} \\
&= y_1^{\alpha_1-1}y_2^{\alpha_2-1}y_3^{\alpha_3-1}\cdots y_{k-1}^{\alpha_{k-1}-1} \\
&\times (1-y_1)^{\sum_{i=2}^k\alpha_i-1}(1-y_2)^{\sum_{i=3}^k\alpha_i-1}(1-y_3)^{\sum_{i=4}^k\alpha_i-1}\cdots(1-y_{k-2})^{\alpha_{k-1}+\alpha_k-1}(1-y_{k-1})^{\alpha_k-1}\,.
\end{aligned}
$$

Note that in the product above we have a natural pairing of the factors $y_i^{\alpha_i-1}$ and $(1-y_i)^{\sum_{j=i+1}^k\alpha_j-1}$ for each value of i in $1\le i\le k-1$. Thus is the multidimensional integral we are attempting to evaluate *decouples* into the product of $k-1$ univariate integrals we have to evaluate $k-1$ of the following integrals

$$
\int_{y_i=0}^1 y_i^{\alpha_i-1}(1-y_i)^{\sum_{j=i+1}^k\alpha_j-1}dy_i\,.
$$

To do this recall the definition of the Beta function $B(a,b)$ given by

$$
B(a,b)\equiv\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,.
$$

Then we see that the integral above is given by

$$
\frac{\Gamma(\alpha_i)\Gamma\left(\sum_{j=i+1}^k\alpha_j\right)}{\Gamma\left(\sum_{j=i}^k\alpha_j\right)}\,.
$$

With the evaluation of these sub integrals the integral of the entire multidimensional integral thus becomes the product of $k-1$ of these results or

$$
\prod_{i=1}^{k-1}\frac{\Gamma(\alpha_i)\Gamma\left(\sum_{j=i+1}^k\alpha_j\right)}{\Gamma\left(\sum_{j=i}^k\alpha_j\right)}=\left(\prod_{i=1}^{k-1}\Gamma(\alpha_i)\right)\left(\prod_{i=1}^{k-1}\frac{\Gamma\left(\sum_{j=i+1}^k\alpha_j\right)}{\Gamma\left(\sum_{j=i}^k\alpha_j\right)}\right)\,.
$$

This second factor above is like a "telescoping series" in that if we write it out we see that a great many terms cancel as follows

$$
\frac{\Gamma\left(\sum_{j=2}^k\alpha_j\right)}{\Gamma\left(\sum_{j=1}^k\alpha_j\right)}\cdot\frac{\Gamma\left(\sum_{j=3}^k\alpha_j\right)}{\Gamma\left(\sum_{j=2}^k\alpha_j\right)}\cdot\frac{\Gamma\left(\sum_{j=4}^k\alpha_j\right)}{\Gamma\left(\sum_{j=3}^k\alpha_j\right)}\cdots\frac{\Gamma\left(\sum_{j=k}^k\alpha_j\right)}{\Gamma\left(\sum_{j=k-1}^k\alpha_j\right)}=\frac{\Gamma\left(\alpha_k\right)}{\Gamma\left(\sum_{j=1}^k\alpha_j\right)}\,.
$$

In summary then we finally find that

$$\int \cdots \int_S \left(\prod_{i=1}^k x_i^{\alpha_i - 1} \right) dx_1 \cdots dx_{k-1} = \frac{\prod_{i=1}^k \Gamma(\alpha_i)}{\Gamma\left(\sum_{j=1}^k \alpha_j\right)},$$

as we were to show.

Warning: I'm not exactly sure how to show that the multidimensional integral above decouples into $k-1$ univariate integrals. Perhaps this would be revealed if one considers a point case where we have a small value for k, say $k=2$. One could then generalize the small value of k procedure to the case of arbitrary k. I have not had time to look into this in more detail. If anyone knows how to show this please contact me.

Exercise 5 (the ratio of X_i to the sum of X_i)

Part (a): We are told that X_i are independent random variables with a gamma distribution with parameters α_i and the same value for β. Lets define Y_i in terms of X_i as

$$Y_i = \frac{X_i}{X_1 + X_2 + \cdots + X_n} \quad \text{for} \quad i = 1, 2, \cdots, n\,.$$

Note that it looks like we have n random variables for $\mathbf{Y}$ but if we introduce the random variable Z defined as

$$Z = \sum_{i=1}^n X_i\,,$$

then given $(X_1, X_2, \cdots, X_n)'$ we can determine all of the variables $(Y_1, Y_2, \cdots, Y_n, Z)'$. In addition, given $(Y_1, Y_2, \cdots, Y_{n-1}, Z)'$ (note no Y_n variable) we can uniquely determine $(X_1, X_2, \cdots, X_n)'$. Thus we will use the p.d.f. of the vector $\mathbf{X}$ to derive the p.d.f of the vector $(Y_1, Y_2, \cdots, Y_{n-1}, Z)'$. We can then integrate out the random variable Z to determine the p.d.f of $(Y_1, Y_2, \cdots, Y_{n-1})'$.

Since X_i's are independent Gamma random variables we have that

$$g_X(x) = \prod_{i=1}^n \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} x_i^{\alpha_i - 1} e^{-\beta x_i}\,.$$

If we define $\alpha_0 = \sum_{i=1}^n \alpha_i$, the p.d.f above becomes

$$g_X(x) = \frac{\beta^{\alpha_0}}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(e^{-\beta \sum_{i=1}^n x_i} \right) \prod_{i=1}^n x_i^{\alpha_i - 1}\,.$$

We will next transform this p.d.f to one over $(Y_1, Y_2, \cdots, Y_{n-1})'$ by using

$$f_{(\mathbf{Y},Z)}(\mathbf{y}, z) = f_{\mathbf{X}}(\mathbf{x}(\mathbf{y}, z)) \left| \frac{\partial \mathbf{x}}{\partial(\mathbf{y}, z)} \right|\,.$$

To evaluate the Jacobian of the above transformation we need the explicit transformation from $(\mathbf{y}, z)$ to $\mathbf{x}$. We have

$$\begin{aligned}
X_i &= ZY_i \quad \text{for} \quad i = 1, 2, \cdots, n-1 \\
X_n &= Z - \sum_{i=1}^{n-1} X_i = Z - Z\sum_{i=1}^{n-1} Y_i = Z\left(1 - \sum_{i=1}^{n-1} Y_i\right) .
\end{aligned}$$

Thus we see that

$$\begin{aligned}
\frac{\partial X_i}{\partial Y_j} &= \begin{cases} 0 & i \neq j \\ Z & i = j \end{cases} \\
\frac{\partial X_i}{\partial Z} &= Y_i \quad \text{for} \quad i = 1, 2, \cdots, n-1 \\
\frac{\partial X_n}{\partial Y_j} &= -ZY_j \\
\frac{\partial X_n}{\partial Z} &= 1 - \sum_{i=1}^{n-1} Y_i .
\end{aligned}$$

Thus we find

$$\frac{\partial \mathbf{x}}{\partial(\mathbf{y}, z)} = \begin{bmatrix}
Z & 0 & 0 & \cdots & 0 & Y_1 \\
0 & Z & 0 & \cdots & 0 & Y_2 \\
0 & 0 & Z & \cdots & 0 & Y_3 \\
\vdots & & & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & Z & Y_{n-1} \\
-ZY_1 & -ZY_2 & \cdots & \cdots & -ZY_{n-1} & 1 - \sum_{i=1}^{n-1} Y_i
\end{bmatrix} .$$

We need to estimate the determinant of this expression, which seems complicated at least I don't see a way to evaluate it that will result in an a-priori simple expression. **Note:** if anyone sees a way to evaluate this determinant simply please let me know. To get around this problem lets see if the expression for $\frac{\partial(\mathbf{y},z)}{\partial \mathbf{x}}$ is any simpler. Maybe it is easier to take the derivative of *that* expression. To evaluate that derivative note that

$$\begin{aligned}
\frac{\partial Y_i}{\partial X_j} &= \frac{\partial}{\partial X_j}\left(\frac{X_j}{Z}\right) = \begin{cases} \frac{1}{Z} & j = i \\ 0 & j \neq i \end{cases} \\
\frac{\partial Z}{\partial X_j} &= 1 \quad \text{for} \quad j = 1, 2, \cdots, n
\end{aligned}$$

Thus in this case we have

$$\frac{\partial(\mathbf{y}, z)}{\partial \mathbf{x}} = \begin{bmatrix}
1/Z & 0 & 0 & \cdots & 0 & 0 \\
0 & 1/Z & 0 & \cdots & 0 & 0 \\
0 & 0 & 1/Z & \cdots & 0 & 0 \\
\vdots & & & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1/Z & 0 \\
1 & 1 & \cdots & \cdots & 1 & 1
\end{bmatrix} .$$

This matrix has a simple derivative, since it is lower triangular and is given by $\left(\frac{1}{Z}\right)^{n-1}$. Thus the determinant we need $\left|\frac{\partial \mathbf{x}}{\partial(\mathbf{y},z)}\right|$ is the reciprocal of this value or Z^{n-1}. Using this derivation we finally find our density function $f_{(\mathbf{Y},Z)}(\mathbf{y},z)$ is given by

$$\begin{aligned} f_{(\mathbf{Y},Z)}(\mathbf{y},z) &= \frac{\beta^\alpha}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(e^{-\beta z}\right) \left(\prod_{i=1}^{n-1} (ZY_i)^{\alpha_i - 1}\right) Z^{\alpha_n - 1} \left(1 - \sum_{i=1}^{n-1} Y_i\right)^{\alpha_n - 1} Z^{n-1} \\ &= \frac{\beta^{\alpha_0} Z^{\alpha_0 - 1} e^{-\beta z}}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(\prod_{i=1}^{n-1} Y_i^{\alpha_i - 1}\right) \left(1 - \sum_{i=1}^{n-1} Y_i\right)^{\alpha_n - 1} . \end{aligned}$$

Next lets integrate out Z to derive the desired p.d.f of just $(Y_1, Y_2, \cdots, Y_{n-1})'$. The expression we have to integrate is given by

$$\int_0^\infty z^{\alpha_0 - 1} e^{-\beta z} dz = \beta^{-\alpha_0} \Gamma(\alpha_0) ,$$

when to evaluate this we make the substitution $v = \beta z$. Thus we find

$$g_{\mathbf{Y}}(\mathbf{y}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(\prod_{i=1}^{n-1} Y_i^{\alpha_i - 1}\right) \left(1 - \sum_{i=1}^{n-1} Y_i\right)^{\alpha_n - 1} ,$$

which is the p.d.f of a Dirichlet random variable as we were to show.

Part (b): Warning: I don't see that this part is any different than Part a except in that here we are only considering the first r elements of X. It seems like even in this case the arguments above will still hold. If someone sees where I am wrong please contact me.

Exercise 6 (order statistics of a uniform random variable are Dirichlet)

One can show [3] that the joint distribution function for the order statistics $\mathbf{Y} = (Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)})'$ when each y_i is drawn from a p.d.f given by $f(y_i)$ is given by

$$f_{Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}}(y_1, y_2, \cdots, y_n) = n! f(y_1) f(y_2) \cdots f(y_n) \quad \text{for} \quad y_1 < y_2 < \cdots < y_n . \tag{72}$$

Then given the transformation from y_i to z_i for $i = 1, 2, \cdots, n$ lets derive the density function of $g_{\mathbf{Z}}(\mathbf{z})$, where $Z = (Z_1, Z_2, \cdots, Z_n)'$. We have that

$$g_{\mathbf{Z}}(\mathbf{z}) = g_{\mathbf{Y}}(\mathbf{y}(\mathbf{z})) \left|\frac{\partial \mathbf{y}}{\partial \mathbf{z}}\right| .$$

Thus to use this formula we need to compute $\frac{\partial \mathbf{y}}{\partial \mathbf{z}}$. From the given expressions we see that

$$\frac{\partial \mathbf{z}}{\partial \mathbf{y}} = \begin{bmatrix} 1 & 0 & & & \\ -1 & 1 & 0 & & \\ 0 & -1 & 1 & \ddots & \\ & \ddots & \ddots & \ddots & 0 \\ & & 0 & -1 & 1 \end{bmatrix} ,$$

which is a matrix that has +1 on its diagonal and −1 on its subdiagonal. The determinate of this matrix is $1^n = 1$. In addition, since the y_i are originally drawn from a uniform distribution where the component densities in Equation 72 are given by $f(y) = 1$ we find that

$$g_{\mathbf{Z}}(\mathbf{z}) = n!\,,$$

Note that this is a Dirichlet process with $\alpha = 1$, since a Dirichlet process over the expanded set of points $(Z_1, Z_2, \cdots, Z_n, Z_{n+1})$ (this last point is a dummy point equal to $Z_{n+1} = 1 - \sum_{i=1}^n Z_n$) when $\alpha = 1$ is given by

$$g_{\mathbf{Z}}(\mathbf{z}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n + \alpha_{n+1})}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_n)\Gamma(\alpha_{n+1})} z_1^{\alpha_1 - 1} z_2^{\alpha_2 - 1} \cdots z_n^{\alpha_n - 1} z_{n+1}^{\alpha_{n+1}-1} = \frac{\Gamma(n+1)}{1^{n+1}} = n!\,,$$

as claimed.

Exercise 7 (what is the mean and covariance of a multidimensional Gaussian)

These two expressions can be derived in a number of ways. One way is by direct integration. For example

$$\begin{aligned}
E(X) &= \int (2\pi)^{-k/2}|\Sigma|^{-1/2}\mathbf{x}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}dx \\
&= (2\pi)^{-k/2}|\Sigma|^{-1/2}\int(\mathbf{x}-\mu+\mu)\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}dx \\
&= (2\pi)^{-k/2}|\Sigma|^{-1/2}\int(\mathbf{x}-\mu)\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}dx \\
&+ \mu(2\pi)^{-k/2}|\Sigma|^{-1/2}\int\exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right\}dx \\
&= 0 + \mu = \mu\,.
\end{aligned}$$

Where the line above last is because of symmetry. An another way to derive this result is to recall that the mean and the covariance can be obtained by evaluating derivatives of the characteristic function at the vector point 0. Since the characteristic function for a multidimensional Gaussian random variable can we written as

$$\zeta(\mathbf{t}) = \exp\left\{i\mathbf{t}'\mu - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right\}\,. \tag{73}$$

Since to expectations in terms of the characteristic function can be computed as

$$E\left(\prod_{j=1}^n x_j^{r_j}\right) = \frac{1}{i^{r_1+r_2+\cdots+r_n}}\left[\frac{\partial^{r_1+r_2+\cdots+r_n}\zeta(\mathbf{t})}{\partial^{r_1}t_1\partial^{r_2}t_2\cdots\partial^{r_n}t_n}\right]\Bigg|_{\mathbf{t}=0} \tag{74}$$

We begin by computing $E(X_j)$ using this method. We find

$$E(X_j) = \frac{1}{i}\left.\frac{\partial\zeta(\mathbf{t})}{\partial t_j}\right|_{\mathbf{t}=0}\,.$$

Since

$$\begin{aligned}\frac{\partial \zeta(t)}{\partial t_j} &= \exp\left\{ it'\mu - \frac{1}{2}t'\Sigma t\right\}\left(i\mu_j - \frac{1}{2}e_j'\Sigma t - \frac{1}{2}t'\Sigma e_j\right) \\ &= \exp\left\{ it'\mu - \frac{1}{2}t'\Sigma t\right\}(i\mu_j - t'\Sigma e_j) \,.\end{aligned}$$

Evaluating this at $t = 0$ gives $E(X_j) = \mu_j$. Next to compute $\text{Cov}(X_i, X_j)$ we need to compute $E(X_iX_j)$. We find

$$\begin{aligned}E(X_iX_j) &= \frac{1}{i^2}\left[\frac{\partial^2\zeta(t)}{\partial t_i \partial t_j}\right]\Bigg|_{t=0} \\ &= -\frac{\partial}{\partial t_j}\left[e^{it'\mu - \frac{1}{2}t'\Sigma t}(i\mu_i - t'\Sigma e_i)\right]\Bigg|_{t=0} \\ &= -\left[e^{it'\mu - \frac{1}{2}t'\Sigma t}(i\mu_j - t'\Sigma e_j)(i\mu_i - t'\Sigma e_i) + e^{it'\mu - \frac{1}{2}t'\Sigma t}(-e_j\Sigma e_i)\right]\Bigg|_{t=0} \\ &= -\left[i\mu_j(i\mu_i) - e_j\Sigma e_i\right] \\ &= \mu_i\mu_j + e_i\Sigma e_j \,.\end{aligned}$$

In the above any i subscript is an index and not the imaginary unit. Now the ijth component of Σ is given by $E(X_iX_j) - E(X_i)E(X_j)$ which can be computed via

$$\Sigma_{ij} = \mu_i\mu_j + e_i\Sigma e_j - \mu_i\mu_j = e_i\Sigma e_j \,,$$

or the ijth element of the matrix Σ as we were to show.

Exercise 8 (the marginal of a multivariate Gaussian is also Gaussian)

To show that the vector $\mathbf{X}_r = (X_1, \cdots, X_r)'$ has a multidimensional normal distribution we use the fact that if X has a multidimensional normal distribution then AX also has a multinomial normal distribution. To use this theorem we pick A so that it selects only the first r components of X. We can do this we take A given by

$$A = \left[\begin{array}{cc} I_{r\times r} & 0_{r\times(k-r)}\end{array}\right] \,,$$

or an $r \times r$ dimensional identity matrix "prepended" to a $r \times (k-r)$ dimensional zero matrix. Then we see that AX is given by

$$AX = \left[\begin{array}{cc} I_{r\times r} & 0_{r\times(k-r)}\end{array}\right]\left[\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_r \\ X_{r+1} \\ \vdots \\ X_k\end{array}\right] = \left[\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_r\end{array}\right] \,,$$

The random variable AX will have a covariance matrix given by $A\Sigma A'$ which in this case will look like

$$\left[\begin{array}{cc} I & 0 \end{array}\right] \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right] \left[\begin{array}{c} I \\ 0 \end{array}\right] = \Sigma_{11} \,.$$

where Σ_{11} is $r \times r$ and is composed of the first r rows and the first r columns of the matrix Σ as claimed.

Exercise 9 (relationships between block components of T and Σ)

We have the matrices Σ and $\Sigma^{-1} = T$ partitioned as

$$\Sigma = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right] \qquad \Sigma^{-1} = T = \left[\begin{array}{cc} T_{11} & T_{12} \\ T_{21} & T_{22} \end{array}\right] . \tag{75}$$

Since we know by definition that $\Sigma^{-1}\Sigma = I$ or

$$\left[\begin{array}{cc} T_{11} & T_{12} \\ T_{21} & T_{22} \end{array}\right] \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right] = \left[\begin{array}{cc} I_{k_1} & 0 \\ 0 & I_{k_2} \end{array}\right] . \tag{76}$$

Next we consider the $(1,2)$ element of the product in the left-hand-side expand the left-hand-side of this product in terms of the blocks above we find

$$T_{11}\Sigma_{12} + T_{12}\Sigma_{22} = 0 \,.$$

Multiplying this equation by T_{11}^{-1} on the left and Σ_{22}^{-1} on the right to get

$$\Sigma_{12}\Sigma_{22}^{-1} + T_{11}^{-1}T_{12} = 0 \,,$$

$$T_{11}^{-1}T_{12} = -\Sigma_{12}\Sigma_{22}^{-1} \,. \tag{77}$$

Next consider the $(1,1)$ component of the product in Equation 76 where we have

$$T_{11}\Sigma_{11} + T_{12}\Sigma_{21} = I_{k_1} \,.$$

Multiplying this equation by T_{11}^{-1} on the left to get

$$\Sigma_{11} + T_{11}^{-1}T_{12}T_{21} = T_{11}^{-1} \,,$$

From Equation 77 we can replace $T_{11}^{-1}T_{12}$ in the above with $-\Sigma_{12}\Sigma_{22}^{-1}$ and get

$$T_{11}^{-1} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \,, \tag{78}$$

which we were to show. Finally, we will consider Σ_{22}^{-1} by looking at the $(2,2)$ element in Equation 76 where we find

$$T_{21}\Sigma_{12} + T_{22}\Sigma_{22} = I_{k_2} \,.$$

Multiplying this equation by Σ_{11}^{-1} on the left we get

$$\Sigma_{22}^{-1} = T_{22} + T_{21}\Sigma_{12}\Sigma_{22}^{-1} \,.$$

From Equation 77 we get that

$$\Sigma_{22}^{-1} = T_{22} - T_{21}T_{11}^{-1}T_{12} \,, \tag{79}$$

another heavily used expression.

Exercise 10 (uncorrelated Gaussian random variables are independent)

The fact that the correlation between X_i and X_j is zero means that

$$E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)'] = 0\,,$$

or that

$$\Sigma_{ij} = 0\,,$$

for $i \neq j$. Thus the p.d.f for the vector $(X_1, X_2, \cdots, X_k)'$ is a multidimensional Gaussian that has a diagonal matrix for is covariance or

$$\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_k^2)\,.$$

With this representation the p.d.f for this vector of random variables has a term in the exponential that looks like

$$(X - \mu)'\Sigma^{-1}(X - \mu) = \sum_{i=1}^{k} \frac{(X_i - \mu_i)}{\sigma_i^2}\,.$$

While the determinant of Σ has an expression given by $|\Sigma| = \prod_{i=1}^{k} \sigma_i^2$. With these simplifications the p.d.f of these random variables looks like

$$g_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{k/2}} \frac{1}{\left(\prod_{i=1}^{k} \sigma_i^2\right)^{1/2}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{k} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \right\}\,,$$

or splitting this into individual factors gives

$$g_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{k} \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left\{ -\frac{1}{2} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \right\}\,.$$

This shows that the joint distribution is written in terms of k products of the marginal distributions and therefore the elements of $\mathbf{X}$ are independent.

Exercise 11 (transformations of independent Gaussian variables are independent)

Since X_i are independent normal random variables each with a variance given by σ^2. When considered as a vector $(X_1, X_2, \cdots, X_k)'$ these are vectors from a multivariate Gaussian distribution with a covariance matrix given by $\Sigma = \sigma^2 I$. The transformed random variables will have a mean given by $A\mu$ and a covariance matrix given by

$$A\Sigma A' = A\sigma^2 I A' = \sigma^2 AA' = \sigma^2 I\,.$$

Since this is a diagonal matrix, using the results from Exercise 10 in this chapter the elements of Y_i are independent and they all have the same variances of σ^2.

Exercise 12 (the random variable $(X - \mu)'\Sigma^{-1}(X - \mu)$ has a χ^2)

We want to show that when X is given by a k dimensional multivariate Gaussian random variable with a mean μ and a covariance matrix Σ that the random variable

$$(X - \mu)'\Sigma^{-1}(X - \mu)\,,$$

has a χ^2 with k degrees of freedom. To see this recall that if we can show that the above expression is the sum of the squares of k independent $N(0, 1)$ Gaussian random variables will have a χ^2 distribution. If we can show that the above expression is of this form we are done. To show this, compute the Cholesky decomposition of the covariance matrix Σ as $\Sigma = GG^T$ and then introduce the random variable Y defined by

$$Y = G^{-1}(X - \mu)\,.$$

Since Y is a linear combination of the vector X it is a Gaussian random vector. In addition, the vector Y has a mean of zero and a covariance given by

$$\text{Cov}(Y) = G^{-1}\text{Cov}(X)G^{-T} = G^{-1}\Sigma G^{-T} = G^{-1}GG^TG^{-T} = I\,.$$

For *Gaussian* random variables the fact that the elements of Y are uncorrelated means that they are independent. Thus the product $Y'Y = \sum_{i=1}^k Y_i^2$ will have a χ^2 distribution with k degrees of freedom. Note that also

$$\begin{aligned} Y'Y &= (X - \mu)'G^{-T}G^{-1}(X - \mu) = (X - \mu)'(GG^T)^{-1}(X - \mu) \\ &= (X - \mu)'\Sigma^{-1}(X - \mu)\,. \end{aligned}$$

Thus the expression $(X - \mu)'\Sigma^{-1}(X - \mu)$ is also distributed as a χ^2 random variable with k degrees of freedom.

Exercise 13 (the distribution of heights)

If we let X be the random variable representing the husbands height and Y be the random variable representing the wife's height. Then we are told that the joint distribution of (X, Y) is given by

$$\begin{aligned} p_{X,Y}(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \\ &\times \exp\left\{-\frac{1}{2\rho(1-\rho)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}\,. \end{aligned}$$

We want to determine if a couple is selected at random what is the probability the husbands height is greater than the wife's. This is given by

$$\begin{aligned} P\{X > Y\} &= \int_{\Omega_{X>Y}} p_{X,Y}(x, y)dxdy \\ &= \int_{y=-\infty}^{+\infty}\int_{x=y}^{\infty} p_{X,Y}(x, y)dxdy\,, \end{aligned}$$

this would be an expression involving the cumulative distribution function of the standard normal.

If we want to know the probability that given wife's height is y what is the probability that the husbands height is greater than y. Thus we want the event $P\{X > Y|Y = y\}$. From the expression for $P_{X,Y}(x, y)$ it can be shown [3], that the conditional distribution $p\{X|Y\}$ is given by

$$p\{X|Y\} = \mathcal{N}\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\right).$$

Thus the probability requested is

$$\begin{aligned} \int_{x=y}^{\infty} p\{X = x|Y = y\}dx &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2(1-\rho^2)}} \\ &\times \int_{x=y}^{\infty} \exp\left\{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[x - \left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)\right)\right]^2\right\}, \end{aligned}$$

which again could be written in terms of the cumulative distribution function for the standard normal.

Exercise 14 (expectations of powers of the determinant)

We would like to evaluate $E(|V|^r)$ when V is given by a Wishart random variable. From the p.d.f for a Wishart random variable we find

$$\begin{aligned} E(|V|^r) &= \int |V|^r f(V|n, \Sigma)dV \\ &= c(n, k)|\Sigma|^{-n/2} \int |V|^r |V|^{(n-k-1)/2} \exp\{-\frac{1}{2}\text{tr}(\Sigma^{-1}V)\}dV \\ &= c(n, k)|\Sigma|^{-n/2} \int |V|^{(n+2r-k-1)/2} \exp\{-\frac{1}{2}\text{tr}(\Sigma^{-1}V)\}dV \end{aligned} \quad (80)$$

where

$$c(n, k) = \left[2^{\frac{nk}{2}}\pi^{\frac{k(k-1)}{4}} \prod_{j=1}^{k} \Gamma\left(\frac{n+1-j}{2}\right)\right]^{-1} \quad (81)$$

Since we know that the Wishart p.d.f is appropriately normalized we can use the definition of $c(n, k)$ above to evaluate integrals like

$$\int_S |V|^{\frac{n-k-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\Sigma^{-1}V)\right\} dV = c(n, k)^{-1}|\Sigma|^{n/2}. \quad (82)$$

Thus to evaluate Equation 80 we can use Equation 82 with n replace with $n + 2r$. Doing this we find

$$E(|V|^r) = c(n, k)|\Sigma|^{-n/2} c(n + 2r, k)^{-1}|\Sigma|^{\frac{n+2r}{2}}.$$

This simplifies and we find

$$
\begin{aligned}
E(|V|^r) &= \frac{c(n,k)}{c(n+2r,k)}|\Sigma|^r = \left(\frac{2^{\frac{(n+2r)k}{2}} \pi^{\frac{k(k-1)}{4}} \prod_{j=1}^k \Gamma\left(\frac{n+2r+1-j}{2}\right)}{2^{\frac{(n)k}{2}} \pi^{\frac{k(k-1)}{4}} \prod_{j=1}^k \Gamma\left(\frac{n+1-j}{2}\right)} \right) |\Sigma|^r \\
&= 2^{rk} \left(\prod_{j=1}^k \frac{\Gamma\left(\frac{n+2r+1-j}{2}\right)}{\Gamma\left(\frac{n+1-j}{2}\right)} \right) |\Sigma|^r \,,
\end{aligned}
$$

which is the result we wanted to show.

Exercise 15 (some expectations involving a Wishart distribution)

Part (a): We want to show that $E(V) = n\Sigma$. To show this recall that when V is distributed as a Wishart distribution then V has the representation given by

$$
V = \sum_{i=1}^n X_i X_i' \,. \tag{83}
$$

Thus taking the expectation of this we find

$$
E(V) = \sum_{i=1}^n E(X_i X_i') = \sum_{i=1}^n \Sigma = n\Sigma \,,
$$

since X_i come from a multivariate normal with a mean of 0 and a covariance matrix of Σ, we can conclude that $E(X_i X_i') = \Sigma$.

Part (b): Since V has a Wishart distribution it has a representation like in Equation 83 so if we look at AVA' we see that

$$
AVA' = \sum_{i=1}^n A X_i X_i' A' = \sum_{i=1}^n (AX_i)(AX_i)' \,. \tag{84}
$$

Since X_i is drawn from a multivariate normal with a mean of 0 and a covariance matrix of Σ, then AX_i are drawn from a multivariate normal random variable with mean 0 and a covariance of $A\Sigma A'$. Thus as $A\Sigma A'$ has the form given in Equation 84, we see that AVA' has a Wishart distribution with n degrees of freedom and a parameter matrix $A\Sigma A'$.

Part (c): We will partition V and Σ as in Equation 75 with V_{11} of size $k_1 \times k_1$, V_{12} of size $k_1 \times k_2$ and similarly for the others. Now to show that V_{11} has a Wishart distribution let

$$
A = \left[\begin{array}{cc} I_{k_1 \times k_1} & 0_{k_1 \times (k-k_1)} \end{array} \right] .
$$

Then we have that

$$
\begin{aligned}
AVA' &= \left[\begin{array}{cc} I & 0 \end{array} \right] \left[\begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array} \right] \left[\begin{array}{c} I \\ 0 \end{array} \right] \\
&= \left[\begin{array}{cc} I & 0 \end{array} \right] \left[\begin{array}{c} V_{11} \\ V_{21} \end{array} \right] = V_{11} \,.
\end{aligned}
$$

Thus as AVA' is given by a Wishart distribution so must be V_{11}. Another way to see this result is to recall that V is given by Equation 83 so that the submatrix V_{11} can be given by a similar sum over the vectors containing only the first k_1 elements of X_i. Since the first k_1 elements of X_i is a Gaussian random vector with mean 0 and covariance Σ_{11} then the matrix V_{11} (since it is the sum of n outer products each of which is given by a Gaussian random variable) is a Wishart random variable with n degrees of freedom and with a parametric matrix Σ_{11}.

Exercise 16 (summing Wishart random variables

Since each V_i are Wishart random variables with n_i degrees of freedom and the same parameter matrix Σ, then V_i has a representation given by Equation 83 but with n replaced by n_i. Thus we see that in this case the V we are given has the representation

$$V = \sum_{i=1}^{r} V_i = \sum_{i=1}^{r} \sum_{j=1}^{n_j} X_j X_j' \,,$$

this later sum is the outer product of $n_1 + n_2 + \cdots + n_r$ terms like $X_j X_j'$, where each X_j is a multidimensional Gaussian random variable with mean 0 and covariance Σ. Since this is the *definition* of a Wishart random variable with $\sum_{i=1}^{r} n_i$ degrees of freedom and a parameter matrix Σ.

Exercise 17 (deriving the characteristic function for a Wishart random variable)

Equation 6 in Section 5.5 is given by Equation 82 with $c(n, k)$ given by Equation 81. Now the characteristic function for a Wishart distribution is defined as

$$\zeta(\mathbf{t}) = E\left[\exp\left(i \sum_{\beta=1}^{k} \sum_{\alpha=1}^{\beta} t_{\alpha\beta} V_{\alpha\beta}\right)\right] \tag{85}$$

We next introduce the T *matrix* (denoted by the t matrix in the book and not to be confused with the precision matrix Σ^{-1} in terms of the elements t_{ij}, introduced above in the expression for the characteristic function, given by

$$T = \begin{bmatrix} 2t_{11} & t_{12} & \cdots & t_{1k} \\ t_{12} & 2t_{22} & \cdots & t_{2k} \\ \vdots & & & \vdots \\ t_{1k} & t_{2k} & \cdots & 2t_{kk} \end{bmatrix} \tag{86}$$

Then the product of the T matrix and the V matrix has an ijth component given by

$$(TV)_{ij} = \sum_{l=1}^{k} T_{il} V_{lj} \,,$$

Thus when $i = j$ we have

$$(TV)_{ii} = \sum_{l=1}^{k} T_{il} V_{li} ,$$

so the trace of the product TV is given by

$$\begin{aligned} \sum_{i=1}^{k} (TV)_{ii} &= \sum_{i=1}^{k} \sum_{l=1}^{k} T_{il} V_{li} \\ &= \sum_{i=1}^{k} T_{ii} V_{ii} + \sum_{i=1}^{l} \sum_{l=1; l \neq i}^{k} T_{il} V_{li} . \end{aligned}$$

Since the iith element of the matrix has a diagonal element of $2t_{ii}$ and since V and T are symmetric we can write the above as

$$\sum_{i=1}^{k} 2t_{ii} V_{ii} + 2 \sum_{i=1}^{k} \sum_{l=1}^{i-1} t_{il} V_{il} .$$

Recalling now the expression in Equation 85 we see that we can write this expression as

$$\zeta(\mathbf{t}) = E \left[\exp \left(i \frac{1}{2} \text{tr}(TV) \right) \right] . \tag{87}$$

Using the definition of the p.d.f. of V to evaluate the above expectation we see that $\zeta(\mathbf{t})$ is given by

$$\begin{aligned} \zeta(\mathbf{t}) &= c(n,k) |\Sigma|^{-n/2} \int_S |V|^{\frac{n-k-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1} V) + \frac{i}{2} \text{tr}(TV) \right\} dV \\ &= c(n,k) |\Sigma|^{-n/2} \int_S |V|^{\frac{n-k-1}{2}} \exp \left\{ -\frac{1}{2} \text{tr}([\Sigma^{-1} - iT] V) \right\} dV . \end{aligned}$$

Using Equation 82 this integral becomes

$$c(n,k) |\Sigma|^{-n/2} c(n,k)^{-1} |(\Sigma^{-1} - iT)^{-1}|^{n/2} = \frac{|\Sigma^{-1}|^{n/2}}{|\Sigma^{-1} - iT|^{n/2}} = \left(\frac{|\Sigma^{-1}|}{|\Sigma^{-1} - iT|} \right)^{n/2} ,$$

as we were to show.

Exercise 18 (the k-dimensional t distribution)

We want to evaluate $E[X]$ for the multidimensional t distribution. Since $X_i = \mu_i + \left(\frac{n}{z}\right)^{1/2} Y_i$ we have that

$$E[X] = E[\mu + \left(\frac{n}{z}\right)^{1/2} Y] = \mu + E\left[\left(\frac{n}{z}\right)^{1/2} Y \right] .$$

Now since Z and Y are independent we can evaluate this later expectation as

$$E\left[\left(\frac{n}{z}\right)^{1/2} Y \right] = \int \left(\frac{n}{z}\right)^{1/2} g_Z(z) dz \cdot \int Y g_{\mathbf{Y}}(\mathbf{y}) dy .$$

The second integral above is zero thus $E[X] = \mu$ as we were to show.

We next want to compute $E[(x-\mu)(x-\mu)']$ for the multidimensional t distribution. Since $X_i - \mu_i = Y_i \left(\frac{n}{z}\right)^{1/2}$ the above is equivalent to $E\left[\left(\frac{n}{z}\right) YY'\right]$. Since Z and Y are independent this expectation can be computed as

$$\begin{aligned} E\left[\left(\frac{n}{z}\right) YY'\right] &= \int\int \left(\frac{n}{z}\right) YY' g_{\mathbf{Y}}(\mathbf{y}) g_Z(z) dy dz \\ &= n \int z^{-1} g_Z(z) dz \cdot \int YY' g_Y(y) dy \,. \end{aligned}$$

We will next evaluate each integral. First we find

$$\begin{aligned} \int z^{-1} g_Z(z) dz &= \left[2^{n/2}\Gamma(n/2)\right]^{-1} \int z^{-1} z^{\left(\frac{n}{2}\right)-1} e^{-z/2} dz \\ &= \left[2^{n/2}\Gamma(n/2)\right]^{-1} \int z^{\frac{n}{2}-2} e^{-z/2} dz \,. \end{aligned}$$

Let $\xi = \frac{z}{2}$ so that $z = 2\xi$ and $d\xi = \frac{dz}{2}$ and the above becomes

$$\begin{aligned} \left[2^{n/2}\Gamma(n/2)\right]^{-1} \int (2\xi)^{\frac{n}{2}-2} e^{-\xi} 2 d\xi &= 2^{-n/2} 2^{n/2-2} \Gamma(n/2)^{-1} \int \xi^{\frac{n}{2}-1-1} e^{-\xi} d\xi \\ &= 2^{-1}\Gamma(n/2)^{-1}\Gamma(\frac{n}{2}-1) \,. \end{aligned}$$

Now recall that $\Gamma(x+1) = x\Gamma(x)$ so that $\Gamma(\frac{n}{2}) = \left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)$ and the above becomes

$$\frac{2^{-1}\Gamma(\frac{n}{2}-1)}{\left(\frac{n}{2}-1\right)\Gamma\left(\frac{n}{2}-1\right)} = \frac{1}{n-2} \,.$$

Thus we have that

$$E\left[\left(\frac{n}{z}\right) YY'\right] = \frac{n}{n-2} \int YY' g_{\mathbf{Y}}(\mathbf{y}) dy \,.$$

From the properties of the multidimensional normal distribution we have that

$$\int YY' g_{\mathbf{Y}}(\mathbf{y}) dy = T^{-1} \,.$$

Thus we conclude that when X has a multidimensional t distribution with parameters n, μ, and T that

$$\mathrm{Cov}(X) = \frac{n}{n-2} T^{-1} \,, \tag{88}$$

as we were to show.

Exercise 19 (the marginal of a t distribution)

If X is a k-dimensional multidimensional t distribution with parameters n, μ, and T if we partition X as $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and T as $T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}$ we want to evaluate

$$f_{X_1}(x_1) = \int f_X(x) dx_2 \,.$$

To do this recall that the density of X or $f_X(x)$ can be written as

$$f_X(x) = \int g_{\mathbf{X},Z}(\mathbf{x}, z)dz = \int_Z g(\mathbf{x}|z)g(z)dz\,.$$

We want to evaluate the marginal distribution of X_1 thus we want to evaluate

$$\begin{aligned} f_{X_1}(x_1) &= \int_{X_2} f_X(\mathbf{x})dx_2 = \int_{X_2} f_{(X_1,X_2)}(\mathbf{x}_1, \mathbf{x}_2)dx_2 \\ &= \int_{X_2}\int_Z g((\mathbf{x}_1, \mathbf{x}_2)|z)g(z)dzdx_2\,. \end{aligned} \tag{89}$$

Note that $g((\mathbf{x}_1, \mathbf{x}_2)|z)$ is a multivariate Gaussian and since $X_i = Y_i\left(\frac{Z}{n}\right)^{-1/2} + \mu_i$ the mean vector for the random variable $(\mathbf{x}_1, \mathbf{x}_2)|z$ is μ and the covariance matrix for this random variable (or $T^{-1}_{X|Z}$ is related to the covariance matrix for Y or T^{-1} as

$$T^{-1}_{X|Z} = \left(\frac{z}{n}\right)^{-1/2} T^{-1}\,.$$

Thus the *precision* matrices are related as

$$T_{X|Z} = \left(\frac{z}{n}\right) T\,. \tag{90}$$

Next changing the order of integration in Equation 89 we have

$$f_{X_1}(x_1) = \int_Z \left[\int_{X_2} g((\mathbf{x}_1, \mathbf{x}_2)|z)dx_2\right] g(z)dz\,.$$

Now from the discussion in Section 5.4 this inner integral is a k_1 multidimensional Gaussian random variable which has a mean μ_1 and a covariance matrix Σ_{11} given by the $(1, 1)$ component of the block partition covariance of

$$T^{-1}_{X|Z} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Here Σ_{11} is $k_1 \times k_1$, Σ_{12} is $k_1 \times k_2$, Σ_{21} is $k_2 \times k_1$, and Σ_{22} is $k_2 \times k_2$. Then the integral expression above becomes

$$\int_{X_2} g((\mathbf{x}_1, \mathbf{x}_2)|z)dx_2 = \frac{1}{(2\pi)^{k_1/2}|\Sigma_{11}|^{1/2}} \exp\left\{-\frac{1}{2}(x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1)\right\}.$$

Now we want to use the results from Section 5.4 to express Σ_{11}^{-1} in terms of the partitioned elements of the precision matrix $T_{X|Z} = \begin{bmatrix} \tilde{T}_{11} & \tilde{T}_{12} \\ \tilde{T}_{21} & \tilde{T}_{22} \end{bmatrix}$. Where $\tilde{T}_{ij}$ has the same dimensions as the matrices Σ_{ij} above. Note also that $\tilde{T}_{ij}$ are the block elements of the precision matrix T for the density $g((\mathbf{x}_1, \mathbf{x}_2)|z)$. To directly use the books result is a bit confusing since in Section 5.4 the book marginalizes out the variable x_1 while in this problem we are marginalizing out x_2. Thus we replace $1 \leftrightarrow 2$ using equation 14 from Section 5.4 we have

$$\Sigma_{11}^{-1} = \tilde{T}_{11} - \tilde{T}_{12}\tilde{T}_{22}^{-1}\tilde{T}_{21}\,.$$

Using Equation 90 to replace the block elements of $\tilde{T}_{ij}$ with the block elements of T which are T_{ij} (without any tildes) we have

$$\Sigma_{11}^{-1} = \frac{z}{n}(T_{11} - T_{12}T_{22}^{-1}T_{21}) .$$

Note that since Σ_{11}^{-1} is of size $k_1 \times k_1$ the determinant of this is given by

$$|\Sigma_{11}^{-1}| = \frac{1}{|\Sigma_{11}|} = \left(\frac{z}{n}\right)^{k_1} |T_{11} - T_{12}T_{22}^{-1}T_{21}| .$$

When we put in $g_Z(z)$ the density of a χ^2 random variable with n degrees of freedom we have $f_{X_1}(x_1)$ given by

$$\begin{aligned}
& \int_Z \left[\frac{1}{(2\pi)^{k_1/2}|\Sigma_{11}|^{1/2}} \exp\left\{ -\frac{1}{2}(x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1) \right\} \right] \left[2^{n/2}\Gamma(n/2) \right]^{-1} z^{\frac{n}{2}-1}e^{-\frac{z}{2}} dz \\
= & \frac{(2\pi)^{-k_1/2}|T_{11} - T_{12}T_{22}^{-1}T_{21}|}{2^{n/2}\Gamma(n/2)n^{k_1/2}} \\
\times & \int_Z z^{\frac{n+k_1}{2}-1} \exp\left\{ -\frac{1}{2}\left[1 + \frac{1}{n}(x_1 - \mu_1)'(T_{11} - T_{12}T_{22}^{-1}T_{21})(x_1 - \mu_1) \right] z \right\} dz \\
= & \frac{(2\pi)^{-k_1/2}|T_{11} - T_{12}T_{22}^{-1}T_{21}|}{2^{n/2}\Gamma(n/2)n^{k_1/2}} \Gamma(\frac{n+k_1}{2}) \\
\times & \left[1 + \frac{1}{n}(x_1 - \mu_1)'(T_{11} - T_{12}T_{22}^{-1}T_{21})(x_1 - \mu_1) \right]^{-\left(\frac{n+k_1}{2}\right)} \left(\frac{1}{2}\right)^{-\left(\frac{n+k_1}{2}\right)} .
\end{aligned}$$

The leading coefficient of this p.d.f above is given by

$$\frac{\Gamma\left(\frac{n+k_1}{2}\right) 2^{n/2}2^{\frac{k_1}{2}}}{(2\pi)^{k_1/2}2^{n/2}\Gamma(n/2)n^{k_1/2}} = \frac{\Gamma(\frac{n+k_1}{2})}{\Gamma(n/2)(n\pi)^{k_1/2}} .$$

Thus we finally end with

$$\begin{aligned}
f_{X_1}(x) &= \frac{\Gamma\left(\frac{n+k_1}{2}\right) |T_{11} - T_{12}T_{22}^{-1}T_{21}|^{1/2}}{\Gamma\left(\frac{n}{2}\right) (n\pi)^{k_1/2}} \\
&\times \left[1 + \frac{1}{n}(x_1 - \mu_1)'(T_{11} - T_{12}T_{22}^{-1}T_{21})(x_1 - \mu_1) \right]^{-\left(\frac{n+k_1}{2}\right)} , \qquad (91)
\end{aligned}$$

This shows that the marginal distribution of X_1 is a k_1-dimensional multivariate t distribution with n degrees of freedom with a location vector μ_1 and a precision matrix of $T_{11} - T_{12}T_{22}^{-1}T_{21}$ as we were to show.

Exercise 20 (the conditional distribution of a multivariate t)

For this problem we want to derive the p.d.f of $f_{X_1}(X_1|X_2 = x_2)$. We can do this by recalling that the conditional distribution is defined as

$$f_{X_1}(X_1|X_2 = x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} .$$

We know from Problem 19 and Equation 91, above that $f_{X_2}(x_2)$ is a k_2-dimensional multi-dimensional t-distribution with n degrees of freedom, a location vector μ_2, and a precision matrix given by $T_{22} - T_{21}T_{11}^{-1}T_{12}$. Note the subscripts here are permuted from those in Equation 91 which is for $f_{X_1}(x_1)$ while here we need $f_{X_2}(x_2)$. Thus this density looks like

$$\begin{aligned} f_{X_2}(x_2) &= \frac{\Gamma(\frac{n+k_2}{2})|T_{22} - T_{21}T_{11}^{-1}T_{12}|^{1/2}}{\Gamma(\frac{n}{2})(n\pi)^{k_2/2}} \\ &\times \left[1 + \frac{1}{n}(x_2 - \mu_2)'(T_{22} - T_{21}T_{11}^{-1}T_{12})(x_2 - \mu_2)\right]^{-\left(\frac{n+k_2}{2}\right)}, \end{aligned}$$

Note that one can show that the precision matrix in the above expression can be written as

$$T_{22} - T_{21}T_{11}^{-1}T_{12} = \Sigma_{22}^{-1},$$

see Exercise 9 on Page 65 above. While the joint density $f_{X_1,X_2}(x_1, x_2)$ looks like

$$f_{X_1,X_2}(x_1, x_2) = \frac{\Gamma(\frac{n+k}{2})|T|^{1/2}}{\Gamma(\frac{n}{2})(n\pi)^{k/2}}\left[1 + \frac{1}{n}(x - \mu)'T(x - \mu)\right]^{-\left(\frac{n+k}{2}\right)}.$$

Writing the quadratic form $(x - \mu)'T(x - \mu)$ in the above expression as in Section 5.4 where it was found that this inner product could be expressed as

$$\begin{aligned} (x - \mu)'T(x - \mu) &= (x_1 - \nu_1)'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}(x_1 - \nu_1) + (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2) \\ &= (x_1 - \nu_1)'T_{11}(x_1 - \nu_1) + (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2), \end{aligned}$$

where ν is the conditional mean of the density $f_{X_1|X_2}(x_1|x_2)$ and is given by

$$\nu_1 = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \mu_1 - T_{11}^{-1}T_{12}(x_2 - \mu_2). \tag{92}$$

Thus putting everything together we have the conditional distribution given by

$$\begin{aligned} f_{X_1}(x_1|X_2 = x_2) &= \frac{\Gamma(\frac{n+k}{2})|T|^{1/2}}{\Gamma(\frac{n}{2})(n\pi)^{k/2}} \cdot \frac{\Gamma(\frac{n}{2})(n\pi)^{k_2/2}}{\Gamma(\frac{n+k_2}{2})|\Sigma_{22}^{-1}|^{1/2}} \\ &\times \frac{\left[1 + \frac{1}{n}(x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2) + \frac{1}{n}(x_1 - \nu_1)T_{11}(x_1 - \nu_1)\right]^{-\left(\frac{n+k}{2}\right)}}{\left[1 + \frac{1}{n}(x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2)\right]^{-\left(\frac{n+k_2}{2}\right)}}. \end{aligned}$$

To simply notation a bit lets define Q_1 and Q_2 such that

$$\begin{aligned} Q_1 &= (x_1 - \nu_1)'T_{11}(x_1 - \nu_1) \\ Q_2 &= (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2), \end{aligned}$$

and recall that by equation 24 in section 5.4 we can write $|T|$ as

$$|T| = |\Sigma^{-1}| = |\Sigma_{22}^{-1}||(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}| = |\Sigma_{22}^{-1}||T_{11}|.$$

Using these expression and by factoring the expression

$$1 + \frac{1}{n}Q_2 + \frac{1}{n}Q_1 = \left[1 + \frac{1}{n}Q_2\right]\left[1 + \frac{\frac{1}{n}Q_1}{1 + \frac{1}{n}Q_2}\right],$$

we can write the above expression for $f_{X_1}(x_1|X_2=x_2)$ as

$$\begin{aligned} f_{X_1}(x_1|X_2=x_2) &= \frac{\Gamma(\frac{n+k}{2})|T_{11}|^{1/2}}{\Gamma(\frac{n+k_2}{2})(n\pi)^{k_1/2}} \\ &\times \left[1+\frac{1}{n}Q_2\right]^{-\left(\frac{n+k}{2}\right)} \left[1+\frac{\frac{1}{n}Q_1}{1+\frac{1}{n}Q_2}\right]^{-\left(\frac{n+k}{2}\right)} \left[1+\frac{1}{n}Q_2\right]^{\left(\frac{n+k_2}{2}\right)} . \end{aligned}$$

Combining the first and the third factors in brackets above we have

$$f_{X_1}(x_1|X_2=x_2) = \frac{\Gamma(\frac{n+k}{2})|T_{11}|^{1/2}}{\Gamma(\frac{n+k_2}{2})(n\pi)^{k_1/2}} \left[1+\frac{1}{n}Q_2\right]^{-\frac{k_1}{2}} \left[1+\frac{\frac{1}{n}Q_1}{1+\frac{1}{n}Q_2}\right]^{-\left(\frac{n+k}{2}\right)} . \quad (93)$$

Now consider just the right most expression in brackets above and note that we can write the negative of the power as

$$\frac{n+k}{2} = \frac{(n+k_2)+k_1}{2} .$$

Thus we need the degrees of freedom of this marginal multivariate t-distribution to be $n+k_2$ not n. With this in mind we get

$$\left[1+\frac{1}{n+k_2}(x_1-\nu_1)'\left[\frac{\frac{n+k_2}{n}T_{11}}{1+\frac{1}{n}Q_2}\right](x_1-\nu_1)\right]^{-\left(\frac{(n+k_2)+k_1}{2}\right)}$$

or

$$\left[1+\frac{1}{n+k_2}(x_1-\nu_1)'\left[\frac{(n+k_2)T_{11}}{n+Q_2}\right](x_1-\nu_1)\right]^{-\left(\frac{(n+k_2)+k_1}{2}\right)} .$$

Thus recalling the definition of Q_2 and Σ_{22}^{-1} from Equation 91 we have that the location vector of the above expression is given by ν_1 or Equation 92. The new precision matrix is given by

$$\begin{aligned} \frac{n+k_2}{n+Q_2}T_{11} &= \frac{n+k_2}{n+(x_2-\mu_2)'\Sigma_{22}^{-1}(x_2-\mu_2)}T_{11} \\ &= \frac{n+k_2}{n+(x_2-\mu_2)'(T_{22}-T_{21}T_{11}^{-1}T_{12})(x_2-\mu_2)}T_{11} . \end{aligned} \quad (94)$$

Thus we could have a multidimensional t distribution with $n+k_2$ degrees of freedom, a location vector ν_1, and a precision matrix given by Equation 94. Then if this is the correct distribution for $f_{X_1}(x_1|X_2=x_2)$ then the *leading coefficient* would need to be given by

$$\frac{\Gamma\left(\frac{n+k_2+k_1}{2}\right)\left|\frac{n+k_2}{n+Q_2}T_{11}\right|}{\Gamma\left(\frac{n+k_2}{2}\right)((n+k_2)\pi)^{k_1/2}} .$$

Since T_{11} is of size $k_1 \times k_1$ this equals

$$\frac{\Gamma\left(\frac{n+k}{2}\right)\left|\frac{n+k_2}{n+Q_2}\right|^{k_1/2}|T_{11}|^{1/2}}{\Gamma\left(\frac{n+k_2}{2}\right)(n+k_2)^{k_1/2}\,\pi^{k_1/2}} = \frac{\Gamma\left(\frac{n+k}{2}\right)|T_{11}|^{1/2}}{\Gamma\left(\frac{n+k_2}{2}\right)(n\pi)^{k_1/2}(1+\frac{1}{n}Q_2)^{k_1/2}} ,$$

which is *exactly* the same as the coefficient in Equation 93 proving the desired result.

Exercise 21 (the distribution of $(X-\mu)'T(X-\mu)/k$)

For this problem we assume that X is a multidimensional t distribution and want to derive the distribution of the expression

$$\frac{1}{k}(X-\mu)'T(X-\mu)$$

By equation 1 of section 5.6 we have that X is given in terms of two variables Y and a Z as

$$X_i - \mu_i = Y_i \left(\frac{Z}{n}\right)^{-1/2} \quad \text{for} \quad i = 1, 2, \ldots k\,. \tag{95}$$

with Y given by a multidimensional normal with zero mean and a precision matrix T and Z given by a χ^2 distribution with n degrees of freedom. In addition, the random variables Y and Z are independent. Thus in terms of Y and Z the above inner product is given by

$$\frac{1}{k}(X-\mu)'T(X-\mu) = \frac{1}{k}\left(\frac{Z}{n}\right)^{-1} Y'TY = \frac{(Y'TY/k)}{(Z/n)}\,.$$

Thus we need to determine the p.d.f of this expression. From the discussion in the book when X is a χ^2 random variable with α degrees of freedom and Y is a χ^2 random variable with β degrees of freedom the variable

$$\frac{(X/\alpha)}{(Y/\beta)}\,,$$

is a F random variable with degrees of freedom α and β. From this since Z is a χ^2 random variable with n degrees of freedom this ratio is given by an F distribution with degrees of freedom k and n if we can show that $Y'TY$ is a χ^2 random variable with k degrees of freedom. Fortunately, in Exercise 36 in Chapter 4 shows that that $Y'TY$ is a χ^2 random variable with k degrees of freedom and the requested result is shown.

Exercise 22 (the distribution of AX when X is a multidimensional t)

If X is a k-dimensional multivariate random variable with n degrees of freedom, a location vector μ and a precision matrix T then it is related to a k-dimensional Gaussian random variable Y with a mean 0 and a precision matrix T and a χ^2 random variable Z with n degrees of freedom as

$$X - \mu = Y\left(\frac{n}{Z}\right)^{1/2} = \frac{Y}{\left(\frac{Z}{n}\right)^{1/2}}\,. \tag{96}$$

Then the U vector defined as AX is related to AY by multiplying Equation 96 by A on the left we have

$$U - A\mu = AY\left(\frac{n}{Z}\right)^{1/2} \tag{97}$$

Thus in this expression we see that as AY and Z are still independent and AY is a m-dimensional random variable with mean 0 and covariance Σ given by

$$\Sigma = AT^{-1}A'\,.$$

Since we assume the product matrix on the right-hand-side of the above is nonsingular the precision matrix of AY is given by $(AT^{-1}A')^{-1}$. Finally, using Equation 97 we have that U is a m-dimensional t-distribution with a location vector $A\mu$, a precision matrix $(AT^{-1}A')^{-1}$, and n degrees of freedom as we were to show.

Exercise 23 (if the joint is a bilateral Pareto so is the marginal)

We are told that the joint p.d.f of X_1 and X_2 is given by a bilateral bivariate Pareto distribution

$$f(x_1, x_2|r_1, r_2, \alpha) = \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}},$$

where $x_1 < r_1$ and $x_2 > r_2$. Then the marginal distribution of x_1 is given by integrating out x_2. Thus we find

$$\begin{aligned} f_{X_1}(x_1|r_1, r_2, \alpha) &= \int_{x_2=r_2}^{\infty} f(x_1, x_2|r_1, r_2, \alpha) dx_2 \\ &= \int_{x_2=r_2}^{\infty} \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}} dx_2 \\ &= \left. \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha (x_2 - x_1)^{-(\alpha+2)+1}}{(-(\alpha+2)+1)} \right|_{r_2}^{\infty} \\ &= \alpha(r_2 - r_1)^\alpha \left[\frac{1}{(r_2 - x_1)^{\alpha+1}} \right], \end{aligned} \quad (98)$$

which is a univariate Pareto distribution over $r_2 - X_1$ with parameters $x_0 \equiv r_2 - r_1$ and α.

The marginal distribution of x_2 is given by the integrating out x_1. Thus we find

$$\begin{aligned} f_{X_2}(x_2|r_1, r_2, \alpha) &= \int_{x_1=-\infty}^{r_1} \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}} dx_1 \\ &= - \left. \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha (x_2 - x_1)^{-\alpha-1}}{(-\alpha-1)} \right|_{-\infty}^{r_1} \\ &= \alpha(r_2 - r_1)^\alpha \left[\frac{1}{(x_2 - r_1)^{\alpha+1}} \right], \end{aligned} \quad (99)$$

which is a univariate Pareto distribution over $X_2 - r_1$ with parameters $x_0 \equiv r_2 - r_1$ and α.

Exercise 24 (the expectation of $(X_2 - X_1)^2$)

Part (a): We are told that the joint distribution of X_1 and X_2 is a bilateral bivariate Pareto distribution

$$f(x_1, x_2|r_1, r_2, \alpha) = \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}},$$

where $x_1 < r_1$ and $x_2 > r_2$. Then we want to evaluate

$$
\begin{aligned}
E[(X_2 - X_1)^2] &= \int_{x_1=-\infty}^{r_1} \int_{x_2=r_2}^{\infty} (x_2 - x_1)^2 f(x_1, x_2 | r_1, r_2, \alpha) dx_2 dx_1 \\
&= \int_{x_1=-\infty}^{r_1} \int_{x_2=r_2}^{\infty} \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^\alpha} dx_2 dx_1 \\
&= \alpha(\alpha+1)(r_2 - r_1)^\alpha \int_{x_1=-\infty}^{r_1} \left. \frac{(x_2 - a_1)^{-\alpha+1}}{(-\alpha+1)} \right|_{x_2=r_2}^{\infty} dx_1 \\
&= \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(-\alpha+1)} \int_{x_1=-\infty}^{r_1} \left[0 - \frac{1}{(r_2 - x_1)^{\alpha-1}} \right] dx_1 \\
&= \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{\alpha - 1} \int_{x_1=-\infty}^{r_1} \frac{dx_1}{(r_2 - x_1)^{\alpha-1}} \\
&= \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{\alpha - 1} \left(- \left. \frac{(r_2 - x_1)^{-\alpha+1+1}}{-\alpha+2} \right|_{x_1=-\infty}^{r_1} \right. \\
&= \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(\alpha-1)(\alpha-2)} \left(\frac{1}{(r_2 - r_1)^{\alpha-2}} \right) \\
&= \frac{\alpha(\alpha+1)(r_2 - r_1)^2}{(\alpha-1)(\alpha-2)} ,
\end{aligned}
$$

as we were to show.

Part (b): To begin consider $E(X_1 X_2)$ when X_1 and X_2 are given by a bivariate Pareto distribution. We find

$$
\begin{aligned}
E(X_1 X_2) &= \int_{x_1=-\infty}^{r_1} \int_{x_2=r_2}^{+\infty} x_1 x_2 \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^\alpha} dx_2 dx_1 \qquad (100) \\
&= \alpha(\alpha+1)(r_2 - r_1)^\alpha \int_{x_1=-\infty}^{r_1} x_1 \int_{x_2=r_1}^{\infty} (x_2 - x_1 + x_1) \frac{1}{(x_2 - x_1)^\alpha} dx_2 dx_1 .
\end{aligned}
$$

This last integral could be split into two parts and each part integrated by hand or integrated using a computer algebra program like Maple or Mathematica designed to help perform these types of manipulations. However, it may in fact be easier to answer this questions using the results from Part (a) above. For example, we have that

$$
E((X_2 - X_1)^2) = E(X_2^2) - 2E(X_1 X_2) + E(X_1^2) = \frac{\alpha(\alpha+1)(r_2 - r_1)^2}{(\alpha-1)(\alpha-2)} .
$$

We can use Equations 54 and 57 to evaluate $E(X_i^2)$ for $i = 1, 2$ and then solve for $E(X_1 X_2)$. Doing this we find

$$
E(X_1 X_2) = \frac{r_1^2 - \alpha r_1 r_2 + r_2^2}{2 - \alpha} . \qquad (101)
$$

Next recall that the correlation between X_1 and X_2 that we are attempting to evaluate can be written in terms of expressions we have evaluated as

$$
\text{Cor}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{E(X_1 X_2) - E(X_1)E(X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} .
$$

When we put all of these expressions together we find this expression equals

$$\text{Cor}(X_1, X_2) = -\frac{1}{\alpha}, \tag{102}$$

as we were to show. Some of the algebra for these problems is worked in the Mathematica notebook `bilateral_Pareto_Derivations.nb`.

Exercise 25 (the limiting behavior of the joint bilateral Pareto)

Warning: I was not able to solve this problem. What follows are some simple notes on the joint bilateral Pareto distribution when we attempt to take the limit of $\alpha \to \infty$. If anyone has any suggestions as to how to do this problem please email me.

We are told that the joint distribution of X_1 and X_2 is a bilateral bivariate Pareto distribution

$$f(x_1, x_2|r_1, r_2, \alpha) = \frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}}.$$

Since for this distribution we have the two facts

$$\begin{aligned} x_1 &< r_1 \quad (103) \\ x_2 &> r_2\,, \quad (104) \end{aligned}$$

by Equation 104 we have that

$$x_2 - x_1 > r_2 - x_1\,,$$

and by Equation 103 we have that the right-hand-side of the above inequality is bounded below as

$$r_2 - x_1 > r_2 - r_1\,.$$

Combining these two inequalities we have

$$x_2 - x_1 > r_2 - r_1 \quad \text{so} \quad \frac{r_2 - r_1}{x_2 - x_1} < 1\,.$$

Now note that the limit as $\alpha \to \infty$ of

$$\frac{\alpha(\alpha+1)(r_2 - r_1)^\alpha}{(x_2 - x_1)^{\alpha+2}},$$

is a limit of the type $\frac{\infty}{\infty}$ and we will need L'Hopital's rule to evaluate it. Writing this limit in the form

$$\lim_{\alpha\to\infty} \frac{\alpha(\alpha+1)}{(x_2 - x_1)^2} \left(\frac{r_2 - r_1}{x_2 - x_1}\right)^\alpha.$$

This is of type $\infty \cdot 0$. We need to take the limit of an expression like

$$\frac{1}{(x_2 - x_1)^2} \lim_{\alpha\to\infty} \frac{\alpha^2 + \alpha}{\xi^\alpha},$$

with $\xi = \frac{x_2 - x_1}{r_2 - r_1} > 1$. To use L'Hopital's rule we need to be able to evaluate $\frac{\xi^\alpha}{d\alpha}$. To do this let $y = \xi^\alpha$ then we see that

$$\ln(y) = \alpha \ln(\xi) \,,$$

so that taking the derivative of this expression w.r.t. α gives

$$\frac{1}{y}\frac{dy}{d\alpha} = \ln(\xi) \,.$$

Now solving for $\frac{dy}{d\alpha}$ we have that

$$\frac{dy}{d\alpha} = y \ln(\xi) = \ln(\xi)\xi^\alpha \,.$$

Thus we find our limit becomes

$$\frac{1}{(x_2 - x_1)^2} \lim_{\alpha \to \infty} \frac{2\alpha + 1}{\ln(\xi)\xi^\alpha} \,,$$

which is still of type $\frac{\infty}{\infty}$. Another application of L'Hopital's rule and the limit of this p.d.f is zero.

Chapter 7 (Utility)

Problem 13 (optimal ordering)

Assume the order is placed for α quarts of drink. Then assuming the demand is for x quarts the profit function will be

$$\text{Profit}(x;\alpha) = \begin{cases} mx - c(\alpha - x) & x < \alpha \\ m\alpha & x > \alpha \end{cases} \tag{105}$$

The the expected profit is given by

$$\begin{aligned} E[\text{Profit}(x;\alpha)] &= \int_0^\alpha (mx - c(\alpha - x)f(x)dx + \int_\alpha^\infty m\alpha f(x)dx & (106) \\ &= \int_0^\alpha (mxf(x)dx - c\alpha \int_0^\alpha f(x)dx + c\int_0^\alpha xf(x)dx + m\alpha \int_\alpha^\infty f(x)dx & (107) \end{aligned}$$

so

$$\frac{dE[\text{Profit}(x;\alpha)]}{d\alpha} = m\alpha f(x) - c\int_0^\alpha f(x)dx - c\alpha f(x) + c\alpha f(x) + m\int_\alpha^\infty f(x)dx + m\alpha(-f(x)) \tag{108}$$

Written in terms of the cumulative distribution function $F(x)$ one has

$$\frac{dE[\text{Profit}(x;\alpha)]}{d\alpha} = -cF(\alpha) + m(1 - F(\alpha)) \tag{109}$$

Setting this expression to zero and solving for $F(\alpha)$ gives

$$F(\alpha) = \frac{m}{c + m} \tag{110}$$

Chapter 8 (decision problems)

Problem 1 (three possible outcomes and three decisions)

In this problem we have four possible outcomes from our experiment w_1,w_2,w_3, and w_4 and three possible decisions. For each decision we compute the expected loss (also called the risk) associated with that decision. From the given table of losses we compute for our specified probability mass function P and an arbitrary decision d

$$\rho(P,d) = \sum_{i=1}^{4} L(w_i,d)P(w_i) \tag{111}$$

Here $P(w_i)$ is the prior probability distribution on the experimental outcomes w_i. Inserting the given probability mass function for our experimental outcomes we obtain

$$\rho(P,d) = \frac{1}{8}L(w_1,d) + \frac{3}{8}L(w_2,d) + \frac{1}{4}L(w_3,d) + \frac{1}{4}L(w_4,d) \tag{112}$$

Now for each of the three decisions we can evaluate this expression. For example for $d = d_1$ we obtain

$$\rho(P, d = d_1) = \frac{1}{8}\cdot 0 + \frac{3}{8}\cdot 1 + \frac{1}{4}\cdot 3 + \frac{1}{4}\cdot 1 = \frac{15}{8} \approx 1.875 \tag{113}$$

The risk for the other decision d_2 and d_3 is computed in the same way. We obtain

$$\rho(P, d = d_2) = \frac{7}{4} \approx 1.75 \tag{114}$$

$$\rho(P, d = d_3) = \frac{9}{8} \approx 1.125 \tag{115}$$

Our final decision is selected by choosing the decision which provides the smallest risk. From the above we see this is decision $d = d_3$ with associated risk $\frac{9}{8}$.

Problem 2 (two possible outcomes and three decisions)

In this problem we have two experimental outcomes w_1 and w_2 and three possible decisions d_1, d_2, and d_3. The loss function is as specified. Since we have only two possible outcomes we can parametrize the probability of each as by $\xi = \Pr(W = w_1)$ with associated $1 - \xi = \Pr(W = w_2)$. Then the expected loss associated with each outcome is given by

$$\rho(\xi,d) = \sum_{i=1}^{2} L(w_i,d)P(w_i) = \xi L(w_1,d) + (1-\xi)L(w_2,d)\,. \tag{116}$$

So for each of the three possible decisions we have that the risk is given by

$$\rho(\xi, d = d_1) = \xi\cdot 0 + (1-\xi)\cdot 8 = 8(1-\xi) \tag{117}$$

$$\rho(\xi, d = d_2) = \xi\cdot 10 + (1-\xi)\cdot 0 = 10\xi \tag{118}$$

$$\rho(\xi, d = d_3) = \xi\cdot 4 + (1-\xi)\cdot 3 = \xi + 3 \tag{119}$$

Now we will have d_3 as the Bayes optimal decision against the distribution W if and only if

$$\xi + 3 < 10\xi \tag{120}$$

giving $\xi > \frac{1}{3}$. In addition, we must have

$$\xi + 3 < 8(1 - \xi) \tag{121}$$

giving $\xi < \frac{5}{9}$. In combination the two conditions give

$$\frac{1}{3} \leq \xi \leq \frac{5}{9} \tag{122}$$

and we have the requested expression.

Problem 3 (continuous outcomes and decisions)

In this case, the experimental outcome is a continuous variable and the possible decisions are also continuous. Thus the expected loss or risk in this case is given by

$$\rho(P, d) = \int_{\Omega_w} L(w, d) dP(w) = \int_0^1 L(w, d) \cdot 2w dw \,. \tag{123}$$

Using the given loss we obtain

$$\begin{aligned}
\rho(P, d) &= 200 \int_0^1 (w - d)^2 w dw & (124)\\
&= 200 \int_0^1 (w^2 - 2wd + d^2) w dw & (125)\\
&= 200 \int_0^1 (w^3 - 2w^2 d + wd^2) dw & (126)\\
&= 200 \left(\frac{w^4}{4} - \frac{2}{3} w^3 d + \frac{d^2 w^2}{2} \right) \Big|_0^1 & (127)\\
&= 200 \left(\frac{1}{4} - \frac{2}{3} d + \frac{d^2}{2} \right) & (128)
\end{aligned}$$

The Bayes optimal decision is the one that minimizes the risk $\rho(P, d)$ with respect to d. Since d can be any any value in the real line we find this minimum by taking the derivative of the above expression. Taking this derivative and setting equal to zero gives

$$\frac{d\rho}{d(d)} = 200 \left(-\frac{2}{3} + d \right) = 0 \tag{129}$$

giving

$$d = \frac{2}{3} \tag{130}$$

in which case the Bayes risk is given by

$$\rho(P, d = \frac{2}{3}) = 200 \left(\frac{1}{4} - \frac{2}{3} \frac{2}{3} + \frac{1}{2} \frac{4}{9} \right) = \frac{50}{9} \tag{131}$$

Problem 4 (the Bayes' decision under different loss functions)

A new loss function $L_0(w, d)$ will yield the same Bayes decision boundary as $L(w, d)$ if it is related by

$$L_0(w, d) = aL(w, d) + \lambda(w) \tag{132}$$

as discussed on Page 125 of the book. In the discrete case given here we can compute L_0 on an individual experimental outcome level. For instance, for the loss function $L(w, d)$ and L_0 we can see if the two are related by a relation like that given in Eq. 132 easily by considering the mapped zero cost $L(w, d)$ element. For instance the elements $L(w_1, d_1)$, $L(w_2, d_2)$, $L(w_3, d_3)$, and $L(w_4, d_3)$. By doing this procedure we obtain

$$\begin{array}{lll} w = w_2 & \lambda(w_1) = +4 & a = 1 \\ w = w_2 & \lambda(w_2) = -1 & a = 1 \\ w = w_3 & \lambda(w_2) = -3 & a = 1 \\ w = w_4 & \lambda(w_4) = -1 & a = 1 \end{array}$$

Since L_0 and L are related as discussed in the text they will yield equivalent Bayes' decision boundaries.

Problem 5 (a convex combination of probability distributions)

By the convexity of the Bayesian risk $\rho^*(P)$ with respect to the probability distribution P we have that

$$\rho^*(\alpha P_1 + (1-\alpha)P_2) \geq \alpha\rho^*(P_1) + (1-\alpha)\rho^*(P_2)\,. \tag{133}$$

As discussed on page 126 of the book. If we can show that

$$\rho^*(\alpha P_1 + (1-\alpha)P_2, d^*) < \alpha\rho^*(P_1, d^*) + (1-\alpha)\rho^*(P_2, d^*)\,. \tag{134}$$

we have the desired equality. To show this inequality we assume that it is not true and derive a contradiction. In that direction, assume that there exists a $d' \neq d^*$ such that

$$\rho(P, d') \leq \rho(P, d^*) \tag{135}$$

i.e. d^* is *not* the Bayes' decision against $P = \alpha P_1 + (1-\alpha)P_2$. Then from the definition of the risk function we have

$$\alpha\rho(P_1, d') + (1-\alpha)\rho(P_2, d') < \alpha\rho(P_1, d^*) + (1-\alpha)\rho(P_2, d^*) \tag{136}$$

or

$$\alpha(\rho(P_1, d') - \rho(P_1, d^*)) + (1-\alpha)(\rho(P_2, d') - \rho(P_2, d^*)) < 0 \tag{137}$$

Since α and $1-\alpha$ are both positive at least one of

$$\rho(P_1, d') - \rho(P_1, d^*) \tag{138}$$

or

$$\rho(P_2, d') - \rho(P_2, d^*) \tag{139}$$

must be negative. This is a contradiction to the fact that d^* is the Bayes' decision against *both* P_1 and P_2 and as such each of Eq. 138 and 139 must be *positive*.

Problem 6 (an incorrect probabilistic specification)

Mathematically this reduces to the following. Compute the difference in the Bayes' risk under the experimental PDF given by $\xi_A(w)$ v.s. that of PDF $\xi_B(w)$ or

$$\rho_A^* - \rho_B^* \tag{140}$$

Which is the additional risk that A will occur due to A's incorrect belief about the experimental distribution of W. In problem 3 (above) we calculated ρ_A^* to be $\frac{50}{9}$, and it remains to calculate the Bayes' risk for the PDF $\xi_B(w)$. As in problem 3 we have for a decision d

$$\begin{aligned}
\rho_B(P,d) &= 100\int_0^1 (w-d)^2 3w^2 dw && (141)\\
&= 300\int_0^1 (w^4 - 2w^3 d + d^2 w^2) dw && (142)\\
&= 300\left(\frac{w^5}{5} - \frac{2}{4}w^4 d + \frac{d^2 w^3}{3}\right)\Big|_0^1 && (143)\\
&= 300\left(\frac{1}{5} - \frac{1}{2}d + \frac{d^2}{3}\right) && (144)
\end{aligned}$$

The Bayes optimal decision is the one that minimizes the risk $\rho_B(P,d)$ with respect to d. Since d can be any any value in the real line we find this minimum by taking the derivative of the above expression. Taking this derivative and setting equal to zero gives

$$\frac{d\rho_B}{d(d)} = -\frac{1}{2} + \frac{2}{3}d = 0 \tag{145}$$

giving

$$d = \frac{3}{4} \tag{146}$$

in which case the Bayes risk is given by

$$\rho_B(P, d = \frac{3}{4}) = \frac{15}{4} \tag{147}$$

so the additional risk incurred by assuming the wrong PDF is given by

$$\rho_A^* - \rho_B^* = \frac{50}{9} - \frac{15}{4} = \frac{65}{36} \tag{148}$$

Problem 7 (a non-unique Bayes' decision)

Computing the risk for each of the 5 available decisions we have (using the definition that $\xi = P(w = w_1)$) we get

$$\begin{aligned}
\rho(P,d_1) &= \xi\cdot 0 + (1-\xi)\cdot 4 = -4\xi + 4 && (149)\\
\rho(P,d_2) &= \xi\cdot 4 + (1-\xi)\cdot 5 = \xi + 5 && (150)\\
\rho(P,d_3) &= \xi\cdot 2 + (1-\xi)\cdot 0 = 2\xi && (151)\\
\rho(P,d_4) &= \xi\cdot 1 + (1-\xi)\cdot 1 = 1 && (152)\\
\rho(P,d_5) &= \xi\cdot 5 + (1-\xi)\cdot 0 = 5\xi && (153)
\end{aligned}$$

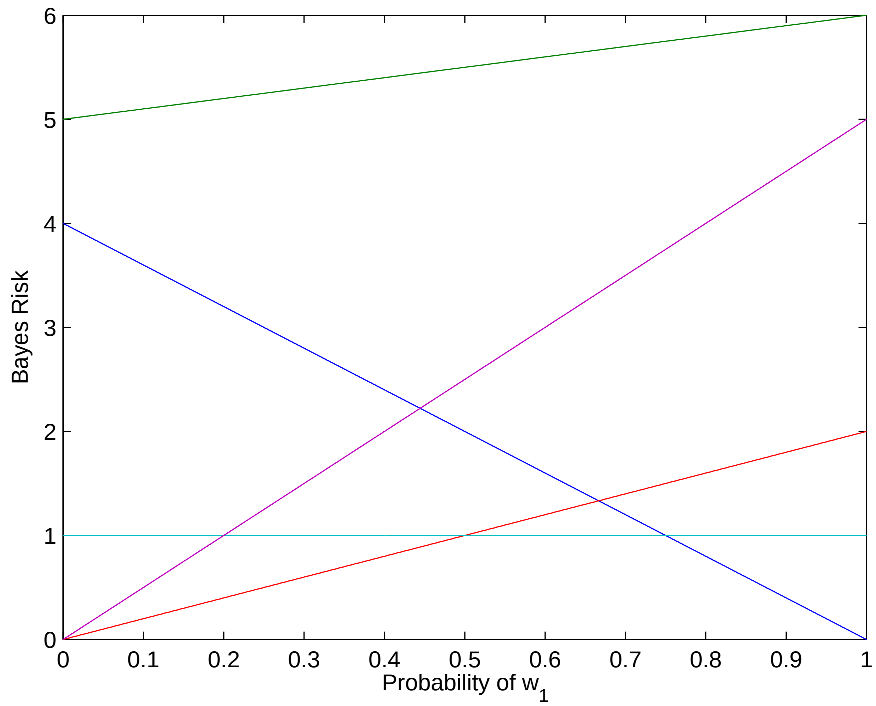

Figure 2: Bayes' risks (for each possible decision d) versus $\xi = P(w = w_1)$ for Problem 7.

Each of the above risks as a function of ξ is shown in figure 2. From this figure one can see that the Bayesian decision is unique *except* at locations where the lowest of two decision functions are equal. From the figure we can see that this occurs at two locations for ξ: ≈ 0.5 and ≈ 0.75. The specific risks that intersect are

$$\rho(\xi, d_3) = \rho(\xi, d_4) \quad \text{or} \quad 2\xi = 1 \quad \Rightarrow \quad \xi = \frac{1}{2} \tag{154}$$

$$\rho(\xi, d_1) = \rho(\xi, d_4) \quad \text{or} \quad -4\xi + 4 = 1 \quad \Rightarrow \quad \xi = \frac{3}{4} \tag{155}$$

Which are the exact values read from the graph. Thus for the two distributions

$$P(w = w_1) = \frac{1}{2} \quad \text{and} \quad P(w = w_2) = \frac{1}{2} \tag{156}$$

$$P(w = w_1) = \frac{3}{4} \quad \text{and} \quad P(w = w_2) = \frac{1}{4} \tag{157}$$

The Bayes' decision is *not* unique.

Problem 8 (all possible Bayes' decisions)

The definition of Bayes risk for the decision d is given by

$$\rho(P, d) = \int_\Omega L(w, d)P(w)dw = \xi L(w = w_1, d) + (1 - \xi)L(w = w_2, d) \tag{158}$$

For each of the given decisions available and the loss specified we have

$$\begin{aligned}
\rho(P, d_1) &= \xi \cdot 1 + (1 - \xi) \cdot 10 = 10 - 9\xi && (159)\\
\rho(P, d_2) &= \xi \cdot 6 + (1 - \xi) \cdot 1 = 5\xi + 1 && (160)\\
\rho(P, d_3) &= \xi \cdot 0 + (1 - \xi) \cdot 13 = 13 - 13\xi && (161)\\
\rho(P, d_4) &= \xi \cdot 2 + (1 - \xi) \cdot 8 = 8 - 6\xi && (162)\\
\rho(P, d_5) &= \xi \cdot 7 + (1 - \xi) \cdot 0 = 7\xi && (163)\\
\rho(P, d_6) &= \xi \cdot 3 + (1 - \xi) \cdot 5 = 5 - 2\xi && (164)\\
\rho(P, d_7) &= \xi \cdot 4 + (1 - \xi) \cdot 4 = 4 && (165)
\end{aligned}$$

Where we have defined $\xi = P(w = w_1)$. Each respective loss functions is plotted in figure 3. For each value of ξ the Bayes' decision is to select the risk that is smallest. For each value of ξ this is easily read from the graph. The decision is *not* unique when two decisions have the same Bayes' risk i.e. $\rho(P, d_i) = \rho(P, d_j)$. From the figure above we see that this when the following risks are equal

$$\begin{aligned}
\rho(\xi, d_5) &= \rho(\xi, d_2) \quad \text{or} \quad 7\xi = 5\xi + 1 \quad \Rightarrow \quad \xi = \frac{1}{2} \approx 0.5 && (166)\\
\rho(\xi, d_2) &= \rho(\xi, d_6) \quad \text{or} \quad 5\xi + 1 = 5 - 2\xi \quad \Rightarrow \quad \xi = \frac{4}{7} \approx 0.57 && (167)\\
\rho(\xi, d_6) &= \rho(\xi, d_1) \quad \text{or} \quad 5 - 2\xi = 10 - 9\xi \quad \Rightarrow \quad \xi = \frac{5}{7} \approx 0.71 && (168)\\
\rho(\xi, d_1) &= \rho(\xi, d_3) \quad \text{or} \quad 10 - 9\xi = 13 - 13\xi \quad \Rightarrow \quad \xi = \frac{3}{4} \approx 0.75 && (169)
\end{aligned}$$

Problem 9 (a problem with no Bayes' decision)

From the definition of the Bayes' risk ρ, for decision d, we have

$$\rho(P, d) = \sum_{w \in \Omega} L(w, d) P(w) \tag{170}$$

For the decision $d = d^*$ this evaluates to

$$\rho(P, d = d^*) = \sum_{w \in \Omega} L(w, d = d^*) P(w) = \frac{1}{2} \sum_{w \in \Omega} P(w) = \frac{1}{2} \,. \tag{171}$$

For the decision $d = d_1$ we have a Bayes' risk of

$$\rho(P, d = d_1) = \sum_{w \in \Omega} L(w, d = d_1) P(w) = \sum_{w \in \Omega \ \setminus \{w_1\}} L(w, d = d_1) P(w) \,. \tag{172}$$

The above can be simplified to

$$1 \sum_{w \in \Omega \ \setminus \{w_1\}} P(w) = 1 (\sum_{w \in \Omega} P(w) - P(w_1)) = 1 - P(w_1) \,, \tag{173}$$

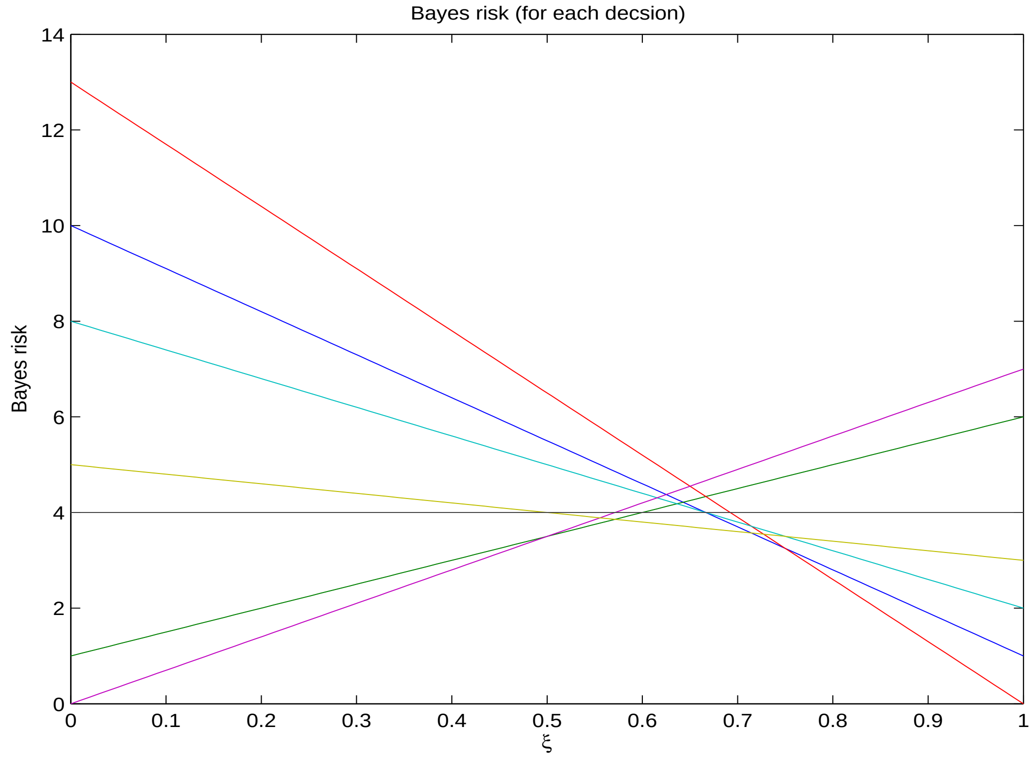

Figure 3: Bayes' risks (for each possible decision d) versus $\xi = P(w = w_1)$ for Problem 8.

since $\sum_{w\in\Omega} P(w) = 1$ by the normalization condition. By similar reasoning for the decision d_2 we have

$$\rho(P, d = d_2) = 1 - P(w_1) - P(w_2)\,. \tag{174}$$

Thus in general, we have for the Bayes' risk for decision d_n is

$$\rho(P, d = d_n) = 1 - \sum_{i=1}^{n} P(w_i)\,. \tag{175}$$

Note that from the above expression we see that the Bayes' risk decreases as n increases or

$$\rho(P, d = d_{n+1}) < \rho(P, d = d_n) \tag{176}$$

and its limiting value is given by

$$\lim_{n\to\infty} \rho(P, d = d_n) = 0\,. \tag{177}$$

Thus d^* cannot be a Bayes' decision since

$$\rho(P, d = d^*) = \frac{1}{2} > \rho(P, d = d_n) \tag{178}$$

for some n large enough by equation 177.

Problem 11 (the Bayes' decision from a measurement)

We begin by assuming that there exists a loss function $L(w, d)$ relating the loss received when the experimental outcome is w and the decision made is d. In this problem we have two decisions d_1 and d_2 two a-priori unknown experimental outcomes w_1 and w_2. The Bayesian formulation of this problem instructs us to compute the expected loss (also called risk) for each possible decision d and select the decision upon which the risk is smallest. The expected loss for making decision d after receiving measurement z is given by

$$\rho(d|z) = \sum_{i=1}^{2} L(w_i, d)p(x, w_i) = \sum_{i=1}^{2} L(w_i, d)p(x|w_i)P(w_i)\,. \tag{179}$$

So in the two decisions case we select action d_1 if

$$\rho(d_1|z) < \rho(d_2|z) \tag{180}$$

or expanding the above summation gives

$$L(w_1, d_1)P(w_1)p(x|w_1) + L(w_2, d_1)P(w_2)p(x|w_2) < \tag{181}$$
$$L(w_1, d_2)P(w_1)p(x|w_1) + L(w_2, d_2)P(w_2)p(x|w_2) \tag{182}$$

Now dividing by the likelihood of class w_1 ($p(x|w_1)$) and defined $P(w_i) = P_i$ and $L_{ij} = L(w_i, d_j)$ we obtain

$$L_{11}P_1 + L_{21}P_2\frac{p(x|w_2)}{p(x|w_1)} < L_{12}P_1 + L_{22}P_2\frac{p(x|w_2)}{p(x|w_1)} \tag{183}$$

Solving for the likelihood ratio

$$\frac{p(x|w_2)}{p(x|w_1)}$$

we obtain the decision region to pick decision d_1 if

$$\frac{p(x|w_2)}{p(x|w_1)}(L_{21}P_2 - L_{22}P_2) < (L_{12}P_1 - L_{11}P_1) \tag{184}$$

Assuming $L_{21} > L_{22}$ which means that it is more costly to make a mistake (the true experimental result is from class 2 while the decision is made assuming the result is from class 1). We can solve for the likelihood ratio giving

$$\frac{p(x|w_2)}{p(x|w_1)} < \frac{(L_{12} - L_{11})P_1}{(L_{21} - L_{22})P_2} \tag{185}$$

The expression encapsulated in Eq. 185 is valid for any likelihood distribution (we have not explicitly required any probabilistic form up to this point) and thus *any* binary decision problem can be started based on this equation.

Because decision problems involving optimal boundaries for two, one-dimensional Gaussian variables are *so* common we will derive the decision boundaries in generality and then specify to the specific parameters given in this problem. In the general problem we assume that the conditional densities for the measured variable z are given by Normal distributions with means μ_i and σ_i in a most general form as

$$
\begin{aligned}
p(z|w_1) &= \mathcal{N}(z;\mu_1,\sigma_1^2) &(186)\\
p(z|w_2) &= \mathcal{N}(z;\mu_2,\sigma_2^2)\,. &(187)
\end{aligned}
$$

Specifically this has the following functional form

$$
p(z|w_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\frac{(z-\mu_i)^2}{\sigma_i^2}}\,. \tag{188}
$$

To continue our derivation specific for this problem we now assume that the given measurement random variable Z is Gaussian distributed. With this assumption we obtain the likelihood ratio of

$$
\frac{p(x|w_2)}{p(x|w_1)} = \frac{\sigma_1}{\sigma_2} exp\left\{-\frac{1}{2}\left[\frac{(z-\mu_2)^2}{\sigma_2^2} - \frac{(z-\mu_1)^2}{\sigma_1^2}\right]\right\} \tag{189}
$$

so Eq. 185 for setting the Bayesian decision boundary becomes

$$
-\frac{1}{2}\left(\frac{(x-\mu_2)^2}{\sigma_2^2} - \frac{(x-\mu_1)^2}{\sigma_1^2}\right) < \log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12}-L_{11})}{(L_{21}-L_{22})}\frac{P_1}{P_2}) \tag{190}
$$

or

$$
\left(\frac{(x-\mu_2)^2}{\sigma_2^2} - \frac{(x-\mu_1)^2}{\sigma_1^2}\right) > -2\log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12}-L_{11})}{(L_{21}-L_{22})}\frac{P_1}{P_2})\,. \tag{191}
$$

Now expanding out each quadratic and grouping terms with the same powers of x we obtain

$$
\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)x^2 - 2\left(\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}\right)x + \frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2} > -2\log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12}-L_{11})}{(L_{21}-L_{22})}\frac{P_1}{P_2}) \tag{192}
$$

For the special case when $\sigma_2 = \sigma_1$ is given below. Without loss of generality we can assume that $\sigma_2 > \sigma_1$ (if this is not true; switch the definition of the classes). This allows us to determine the sign of the coefficient of x^2. We easily conclude that

$$
\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} < 0\,.
$$

We can divide by the difference above giving a more pure quadratic equation

$$
x^2 - 2\frac{\frac{\mu_2}{\sigma_2^2} - \frac{\mu_1}{\sigma_1^2}}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}}x + \frac{\frac{\mu_2^2}{\sigma_2^2} - \frac{\mu_1^2}{\sigma_1^2}}{\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}} < -\frac{2}{\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right)}\log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12}-L_{11})}{(L_{21}-L_{22})}\frac{P_1}{P_2})\,. \tag{193}
$$

Multiplying by $\sigma_1^2\sigma_2^2$ on the top and bottom of each fraction gives

$$
x^2 + 2\left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)x + \left(\frac{\mu_1^2\sigma_2^2 - \mu_2^2\sigma_1^2}{\sigma_2^2 - \sigma_1^2}\right) < \frac{2\sigma_1^2\sigma_2^2}{(\sigma_2^2 - \sigma_1^2)}\log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12}-L_{11})}{(L_{21}-L_{22})}\frac{P_1}{P_2})\,. \tag{194}
$$

To complete the square of the above equation add the square of one half of the coefficient of the x term to both sides or

$$
\left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^2 \tag{195}
$$

This gives

$$\left(x + \left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)\right)^2 + \left(\frac{\mu_1^2\sigma_2^2 - \mu_2^2\sigma_1^2}{\sigma_2^2 - \sigma_1^2}\right) < \quad (196)$$

$$\frac{2\sigma_1^2\sigma_2^2}{(\sigma_2^2 - \sigma_1^2)} \log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) + \left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^2 . \quad (197)$$

or

$$\left(x + \left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)\right)^2 < \quad (198)$$

$$\frac{2\sigma_1^2\sigma_2^2}{(\sigma_2^2 - \sigma_1^2)} \log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) + \left(\frac{\mu_1^2\sigma_2^2 - \mu_2^2\sigma_1^2}{\sigma_2^2 - \sigma_1^2}\right) + \left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^2 . \quad (199)$$

To simplify notation we will define $\mathcal{R}$ and $\mathcal{M}$ both functions of $(\mu_1, \sigma_1, \mu_2, \sigma_2, P_1, P_2, L)$ as

$$\mathcal{R} = \frac{2\sigma_1^2\sigma_2^2}{(\sigma_2^2 - \sigma_1^2)} \log(\frac{\sigma_2}{\sigma_1}\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) + \left(\frac{\mu_1^2\sigma_2^2 - \mu_2^2\sigma_1^2}{\sigma_2^2 - \sigma_1^2}\right) + \left(\frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2}\right)^2 \quad (200)$$

$$\mathcal{M} = \frac{\mu_2\sigma_1^2 - \mu_1\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \quad (201)$$

Thus our decision boundary becomes: decide d_1 when

$$(x + \mathcal{M})^2 < \mathcal{R} \quad (202)$$

or

$$-\sqrt{\mathcal{R}} - \mathcal{M} < x < +\sqrt{\mathcal{R}} - \mathcal{M} \quad (203)$$

One would make the decision d_2 if this result did not hold.

We now consider a few special cases of the above general relationship.

Equal variances: $\sigma_1 = \sigma_2 = \sigma$

In this case equation 192 simplifies and becomes

$$-2\left(\frac{\mu_2 - \mu_1}{\sigma^2}\right)x + \frac{\mu_2^2 - \mu_1^2}{\sigma^2} > -2\log(\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) \quad (204)$$

or

$$-2(\mu_2 - \mu_1)\left(x - \frac{\mu_2 + \mu_1}{2}\right) > -2\sigma^2\log(\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) \quad (205)$$

or

$$(\mu_2 - \mu_1)\left(x - \frac{\mu_2 + \mu_1}{2}\right) < \sigma^2\log(\frac{(L_{12} - L_{11})}{(L_{21} - L_{22})}\frac{P_1}{P_2}) \quad (206)$$

To simplify further we must assume something about the sign of $\mu_2 - \mu_1$.

Equal means: $\mu_1 = \mu_2 = \mu$

Then Eq. 192 gives

$$\left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2}\right) x^2 > -2 \log \frac{\sigma_2}{\sigma_1} \frac{(L_{12} - L_{11})}{(L_{21} - L_{22})} \tag{207}$$

FINISH!!!

Chapter 9 (Conjugate Prior Distributions)

Problem 23 (number of samples required for a given confidence)

We know that $p_X(x)$ is given as $N(\mu, 4)$ and lets assume a mean for our prior of μ_0. Then from the information given the prior is expressed as $p(\mu)$ $N(\mu_0, 9)$. From the class notes after n samples have been received from a PDF $p_X(x)$ and Bayesian learning is performed then the posteriori distribution of μ is given by

$$p(\mu|D) = N(\mu_n, \sigma_n^2) \tag{208}$$

with

$$\mu_n = (\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2})\hat{\mu}_n + (\frac{\sigma^2}{n\sigma_0^2 + \sigma^2})\mu_0 \tag{209}$$

$$\sigma_n^2 = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2} \tag{210}$$

Here $\hat{\mu}_n = \frac{1}{n}\sum_i x_i$ is the sample mean. This is also equation number 34 and 35 on page 94 in [2]. In this problem $\sigma_0^2 = 9$, $\sigma^2 = 4$, so the above simplifies to

$$\mu_n = (\frac{9n}{9n + 4})\hat{\mu}_n + (\frac{4}{9n + 4})\mu_0 \tag{211}$$

$$\sigma_n^2 = \frac{9 * 4}{9n + 4} = \frac{36}{9n + 4} \tag{212}$$

A 95 percent confidence interval will lie between 1.96 standard deviations of the mean μ_n. If we desire this interval to be of length 1 (or smaller) then we must have

$$(\mu_n + 1.96 * \sigma_n) - (\mu_n - 1.96 * \sigma_n) = 2 * 1.96 * \sigma_n = 1 \tag{213}$$

Written in terms of the number of samples this is

$$\sigma_n^2 = \frac{36}{9n + 4} = (\frac{1}{2 * 1.96})^2 \tag{214}$$

Solving this for n gives $n \approx 61.02$. Since n must be an integer one should take $n \geq 62$.

References

[1] W. E. Boyce and R. C. DiPrima. *Calculus.*

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* Wiley, 2001.

[3] S. Ross. *A First Course in Probability.* Macmillan, 3rd edition, 1988.

[4] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Third Edition.* Academic Press, Inc., Orlando, FL, USA, 2006.