Main Examination period 2019

MTH5126: Statistics for Insurance

Duration: 2 hours

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

You should attempt ALL questions. Marks available are shown next to the questions.

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

The New Cambridge Statistical Tables 2nd Edition are provided.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any unauthorised notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: D. Boland, S. Liverani

Question 1. [6 marks] An urn B_1 contains 2 white and 3 black balls. A second urn B_2 contains 3 white and 4 black balls. One urn is selected at random and a ball is drawn from it. If the ball drawn is black, find the probability that the urn chosen was B_1. It should be assumed that each ball in the selected urn is drawn with equal probability. **[6]**

Let E_1, E_2 denote the events of selecting B_1, B_2 respectively.
Then:

$$P(E_1) = P(E_2) = \frac{1}{2}$$

[1]

Let B denote the event that the ball chosen from the selected urn was black.
Then we have to find $P(E_1|B)$ **[1]**
We have: $P(B|E_1) = \frac{3}{5}$ and $P(B|E_2) = \frac{4}{7}$ **[1]**
So, by Bayes' Theorem we get:

$$P(E_1|B) = \frac{P(B|E_1)P(E_1)}{P(B|E_1)P(E_1) + P(B|E_2)P(E_2)}$$

[1]

$$= \frac{\frac{1}{2} \times \frac{3}{5}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{4}{7}}$$
$$= \frac{21}{41}$$

[2]

Question 2. [4 marks]
Consider a general game:

		Player A	
		I	**II**
Player B	**1**	a	b
	2	d	c

(a) Define a **saddle point.** **[2]**

(b) State the conditions for which a saddle point exists in this game. **[2]**

(a) A pair of strategies are in equilibrium if and only if the element $L(a,b)$ corresponding to the equilibrium is both the largest in its column and the smallest in its row. Such equilibrium is called a *saddle point*.

(b) A saddle point is the largest element in its column and the smallest in its row.

We have 4 possibilities for a saddle: a, b, c, d. Assuming these are equally likely, then the conditions for a (for example) would be:

$$a > d$$
$$a < b$$

Question 3. [15 marks] The table below shows claims paid on a portfolio of general insurance policies. You may assume that claims are fully run off after three years.

Underwriting year	Development Year			
	0	1	2	3
2012	650	412	217	91
2013	703	489	262	
2014	711	456		
2015	678			

Past claims inflation has been 2% p.a. However, it is expected that future claims inflation will be 4% p.a. Use the inflation adjusted chain ladder method to calculate the outstanding claims on the portfolio. [**15**]

We begin by finding the claim amounts uplifted to 2015 prices.

Underwriting year	Development Year			
	0	1	2	3
2012	689.79	428.64	221.34	91
2013	731.40	498.78	262	
2014	725.22	456		
2015	678			

[**3**]

Accumulating gives

Underwriting year	Development Year			
	0	1	2	3
2012	689.79	1118.43	1339.77	1430.77
2013	731.40	1230.18	1492.18	
2014	725.22	1181.22		
2015	678			

[**2**]

Hence the development factors are

$$DF_{0,1} = \frac{1118.43 + 1230.18 + 1181.22}{689.79 + 731.40 + 725.22} = \frac{3529.83}{2146.41} = 1.644527$$

$$DF_{1,2} = \frac{1339.77 + 1492.18}{1118.43 + 1230.18} = \frac{2831.95}{2348.61} = 1.205798$$

$$DF_{2,3} = \frac{1430.77}{1339.77} = 1.067922$$

[**3**]

The completed table at 2015 prices is

Underwriting year	Development Year			
	0	1	2	3
2012				
2013				1593.53
2014			1424.31	1521.05
2015		1114.99	1344.45	1435.77

[**2**]

Differencing gives

Underwriting year	Development Year			
	0	1	2	3
2012				
2013				101.35
2014			243.09	96.74
2015		436.99	229.46	91.32

[**1**]

Applying inflation for future years gives

Underwriting year	Development Year			
	0	1	2	3
2012				
2013				105.40
2014			252.81	104.63
2015		454.47	248.18	102.72

[**3**]

So the outstanding claims are

$$105.40 + 252.81 + \cdots + 102.72 = 1268.21$$

[**1**]

Question 4. [12 marks]
A decision-maker with decreasing risk-aversion has assets of £30 and has a decision problem with the following structure:

	Outcome	
Decision	θ_1	θ_2
d_1	-10	$+5$
d_2	$+15$	-5
Probability	0.3	0.7

The entries in the table represent gains or losses in pounds. For example, with decision d_1 and outcome θ_1 the decision-maker will end up with assets of £20 and with decision d_2 and outcome θ_2 they will end up with assets of £25.
You are given the following excerpt from the decision-maker's utility table:

£x	$u(x)$	£x	$u(x)$
0	0.000	180	0.797
10	0.221	185	0.802
15	0.300	190	0.807
20	0.364	195	0.811
25	0.415	200	0.816
30	0.458	210	0.825
35	0.493	220	0.834
40	0.523	230	0.842
45	0.548	240	0.849
50	0.570	250	0.847

(a) Advise the decision-maker on whether decision d_1 or d_2 is the better course of action. [**7**]

(b) Would your advice remain the same if the decision-maker's assets were £200? [**5**]

You should justify any advice given and show clear workings.

(a) The investor currently has £30.
The first thing we should check is the monetary amounts he would have with each decision.
d_1 and θ_1 means he would lose £10, leaving him with £20.
d_1 and θ_2 means he would win £5, leaving him with £35.
d_2 and θ_1 means he would win £15, leaving him with £45.
d_2 and θ_2 means he would lose £5, leaving him with £25. **[2]**

We can now use the tables provided to check the relevant associated utilities.

$$u(30) = 0.458$$
$$u(20) = 0.364$$
$$u(35) = 0.493$$
$$u(45) = 0.548$$
$$u(25) = 0.415$$

[2]

Finally we can calculate the expected utilities and check which decision is better on this criteria.

Decision d_1: $\bar{u}(x) = 0.3 \times 0.364 + 0.7 \times 0.493 = 0.4543$ **[2]**
Decision d_2: $\bar{u}(x) = 0.3 \times 0.548 + 0.7 \times 0.415 = 0.4549$ **[2]**
Therefore we prefer decision d_2 as the expected utility is higher at 0.4549. **[2]**
It is worth noting that a further decision, say d_0 corresponding to not investing at all would have a higher utility than d_1 and d_2 at 0.458.
Should a student point this out then discretionary marks will be awarded.

(b) When the decision-maker has assets of £200 then the expected utilities are:

$$\begin{aligned} u(200) &= 0.816 \\ u(190) &= 0.807 \\ u(205) &= (0.816 + 0.825)/2 = 0.8205 \\ u(215) &= (0.825 + 0.834)/2 = 0.8295 \\ u(195) &= 0.811 \end{aligned}$$

[2]

Decision d_1: $\bar{u}(x) = 0.3 \times 0.807 + 0.7 \times 0.8205 = 0.81645$
Decision d_2: $\bar{u}(x) = 0.3 \times 0.8295 + 0.7 \times 0.811 = 0.81655$

So we prefer decision d_2 again and in fact the utility is very close to the initial amount invested. **[3]**

Question 5. [20 marks]

(a) A new insurance policy is sold to a limited number of existing policy holders. The number of claims in the first six months is 18. Suppose the number of claims each month is assumed to have a Poisson distribution with mean θ, independent of all other months. Write down the likelihood function. **[4]**

(b) The prior distribution of θ is given by a Gamma distribution. When designing the policy it was thought that based on the claims history of these policy holders the mean number of claims per month would be 4 claims with variance 1. Show that a Gamma distribution $Gamma(16,4)$ is a suitable prior. **[2]**

(c) Find the posterior distribution of θ. **[4]**

(d) Find the form of the Bayes estimate of θ under squared error loss. **[6]**

(e) Calculate the value of the Bayes estimate. Show it can be written as a weighted average of the prior mean and the data mean and interpret the weights. **[4]**

(a) The likelihood is given by

$$p(\underline{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta) \propto e^{-n\theta}\theta^{S}$$

[2]

where $S = \sum x_i$. **[1]**

Here $n = 6$ and $S = 18$ so the likelihood is proportional to $e^{-6\theta}\theta^{18}$ **[1]**

(b) The mean and variance of a $Ga(a,b)$ distribution are a/b and a/b^2 (from table of distributions). So $Ga(16,4)$ has mean 4 and variance 1. **[2]**

(c) The prior $p(\theta) \propto \theta^{15}e^{-4\theta}$ **[1]**

So the posterior $p(\theta|(\underline{x})) \propto \theta^{15}e^{-4\theta}\theta^{18}e^{-6\theta} \propto \theta^{33}e^{-10\theta}$ **[1]**

So the posterior distribution is $Ga(34,10)$. **[2]**

(d) If we have squared error loss $l(t,\theta) = (t-\theta)^2$ then the posterior mean is the Bayes estimate. We can show this as follows. **[1]**
We wish to minimise the expected loss

$$\min_t \int (t-\theta)^2 p(\theta|\underline{x})d\theta.$$

[1]

Differentiating with respect to t we have

$$\int 2(t-\theta)p(\theta|(\underline{x}))d\theta$$

[1]

setting this equal to zero

$$\hat{t}\int p(\theta|(\underline{x}))d\theta = \int \theta p(\theta|(\underline{x}))d\theta$$

[1]

and since the integral of the posterior density is one we have

$$\hat{t} = \int \theta p(\theta|(\underline{x}))d\theta = E[\theta|(\underline{x})].$$

[2]

(e) We have established in part (c) that the posterior distribution is $Ga(34,10)$, so it follows that the posterior mean is 3.4 and the posterior variance is 0.34. The posterior mean 3.4 is the Bayes estimate. **[1]**
The prior mean was 4 and the sample mean $18/6 = 3$. **[1]**
The weights are the relative information provided by the data (6 observations) and the prior (equivalent to 4 observations). **[1]**

We can write the posterior mean as a weighted average of the sample mean and prior mean as follows:

$$3.4 = 3\times\frac{6}{10} + 4\times\frac{4}{10}$$

[1]

Question 6. [10 marks] The table below shows aggregate annual claim statistics for three risks over a period of eight years. Annual aggregate claims for risk i in year j are denoted by X_{ij}.

Risk i	$\bar{X}_{i.} = \frac{1}{8}\sum_{j=1}^{8} X_{ij}$	$S_i^2 = \frac{1}{7}\sum_{j=1}^{8}(X_{ij} - \bar{X}_{i.})^2$
1	213.11	411.19
2	91.15	94.23
3	134.23	38.6

(a) What are the assumptions of the Empirical Bayes Credibility Theory (EBCT) Model 1? [**2**]

(b) Calculate the credibility premium of each risk under the assumptions of EBCT Model 1. [**8**]

(a) The assumptions for EBCT Model 1

(i) The distribution of each X_j depends on a parameter, denoted θ,whose value is fixed (and the same for all the X_j's) but is unknown.

(ii) Given θ, the X_j's are independent and identically distributed.

(b) The overall mean is

$$\frac{213.11 + 91.15 + 134.23}{3} = 146.16$$

[1]

and

$$E(S^2(\theta)) = \frac{1}{3}\sum_i S_i^2 = \frac{411.19 + 94.23 + 38.6}{3} = 181.34$$

[1]

$$\begin{aligned} Var(m(\theta)) &= \frac{1}{2}\sum (X_i - \bar{X})^2 - \frac{1}{8}E(S^2(\theta)) \\ &= \frac{1}{2}\left[(213.11 - 146.16)^2 + (91.15 - 146.16)^2 + (134.23 - 146.16)^2\right] - \frac{181.34}{8} \\ &= \frac{1}{2}(4481.86 + 3026.47 + 142.40) - 22.6675 \\ &= 3825.36 - 22.6675 = 3802.70 \end{aligned}$$

[2]

So the credibility factor is

$$Z = \frac{8}{8 + \frac{181.34}{3802.70}} = 0.994074$$

[1]

Therefore the credibility premia for the risks are

$$0.994074 \times 213.11 + (1 - 0.994074) \times 146.16 = 212.71$$

[1]

$$0.994074 \times 91.15 + (1 - 0.994074) \times 146.16 = 91.48$$

[1]

$$0.994074 \times 134.23 + (1 - 0.994074) \times 146.16 = 134.30$$

[1]

Question 7. [10 marks] An insurance company has a portfolio of policies under which individual loss amounts follow an exponential distribution with mean λ^{-1}. In one year, the insurer observes 75 claims with mean claim amount 107.8.

(a) Show that the maximum likelihood estimate of λ is 0.0093. **[8]**

(b) Find the probability that an individual loss amount will be greater than 107.8. **[2]**

(a) The likelihood function for the data is proportional to

$$\left(\lambda e^{-107.8\lambda}\right)^{75}$$

[3]

so the log-likelihood is

$$C + 75\log\lambda - 75\times 107.8\lambda.$$

[1]

Hence

$$\frac{d\log L}{d\lambda} = \frac{75}{\lambda} - 75\times 107.8.$$

[1]

Setting this equal to zero and solving we see that the maximum likelihood estimate of λ is

$$\hat{\lambda} = \frac{75}{75\times 107.8} = \frac{1}{107.8} = 0.0093.$$

[2]

This is a maximum as

$$\frac{d^2\log L}{d\lambda^2} = -\frac{75}{\lambda^2} < 0.$$

[1]

(b) The probability of an individual loss amount exceeding 107.8 is

$$\int_{107.8}^{\infty} \lambda e^{-\lambda x}dx = \left[-e^{\lambda x}\right]_{107.8}^{\infty} = e^{-107.8\lambda}$$

Using $\hat{\lambda} = 1/107.8$, the probability is

$$e^{-107.8\times\frac{1}{107.8}} = e^{-1}.$$

[2]

Question 8. [15 marks]
A manufacturing company is analysing the number of accidents that occur each year on the factory floor. They believe that the number of accidents per year, N, has a geometric distribution with parameter 0.7 so that:

$$P(N = n) = 0.7 \times 0.3^n, \qquad n = 0, 1, 2, ..$$

For each accident i, the number of employees injured is Y_i, where $Y_i = X_i + 1$ and X_i is believed to have a $Poisson(1.8)$ distribution.

The company has taken out an insurance policy, which provides a benefit of £5,000 to each injured employee, up to a maximum of three employees per accident, irrespective of the level of injury. There is no limit on the number of accidents that may be claimed for in a year.

(a) Show that $E(S) = 1.0166$ and $var(S) = 3.6859$, where S is the total number of employees claiming benefit in a year under this policy. [**9**]

(b) Hence find the mean and variance of the aggregate amount paid out under this policy in a year. [**6**]

(a) The aggregate amount paid out by the company is $5,000S$, where:

$$S = Z_1 + \ldots + Z_n$$

is the total number of employees claiming benefit in a year. **[1]**
N has the distribution given in the question and:

$$Z_i = \begin{cases} Y_i & Y_i < 3 \\ 3 & Y_i \geq 3 \end{cases}$$

where $Y_i = X_i + 1$ and X_i has a $Poisson(1.8)$ distribution. We note that:
$P(Y_i = 1) = e^{-1.8}$ (The Poisson probability that $X_i = 0$)
$P(Y_i = 2) = 1.8e^{-1.8}$ (The Poisson probability that $X_i = 1$)
$P(Y_i \geq 3) = 1 - e^{-1.8} - 1.8.e^{-1.8} = 1 - 2.8e^{-1.8}$ **[3]**

We must first find the mean and variance of Z_i:

$$E(Z_i) = \sum_{z=1}^{2} zP(Y_i = z) + 3P(Y_i \geqslant 3) = e^{-1.8} + 3.6e^{-1.8} + 3(1 - 2.8e^{-1.8}) = 3 - 3.8e^{-1.8} = 2.37186$$

[3]

$$E(Z_i^2) = \sum_{z=1}^{2} z^2 P(Y_i = z) + 3^2 P(Y_i \geqslant 3) = e^{-1.8} + 7.2e^{-1.8} + 9(1 - 2.8e^{-1.8}) = 9 - 17e^{-1.8} = 6.1899$$

$$var(Z_i) = E(Z_i^2) - (E(Z_i))^2 = 6.1899 - 2.37186^2 = 0.5642$$

[2]

The expectation and the variance of the number of claims per year are as follows.

$$E(N) = \frac{1-p}{p} = \frac{0.3}{0.7} = 0.4286$$

[1]

$$var(N) = \frac{1-p}{p^2} = \frac{0.3}{0.7^2} = 0.6122$$

[1]

So the mean and variance of S are:

$$E(S) = E(Z_i)E(N) = 2.37186 \times 0.4286 = 1.0166$$

[1]

$$var(S) = (E(Z_i))^2 var(N) + var(Z)E(N) = 2.37186^2 \times 0.6122 + 0.5642 \times 0.4286 = 3.6859$$

[2]

(b) The mean and variance of $5,000S$ are:

$$E(5,000S) = 5,000E(S) = 5,082.5$$

[3]

$$var(5,000S) = 5,000^2 var(S) = 92,152,500$$

[3]

[Follow through marks were awarded where a calculation error was made]

Question 9. [8 marks]

(a) Explain the difference between **IBNR** and an **Outstanding Claims Reserve.** **[4]**

(b) Explain the difference between **Proportional** and **Non-Proportional** reinsurance and give two examples of each. **[4]**

(a) An IBNR reserve is required in respect of claims that have been incurred but not reported. **[1]**
This means that the claim event has occurred but the claim has not yet been reported to the insurer. **[1]**
An Outstanding Claims reserve is required in respect of claims that may have been reported but have not yet been closed. **[1]**
OCR's are usually calculated by run-off triangle method whereas IBNR uses Loss Ratios **[1]**
IBNR happens earlier in the typical four part claims process Claim event occurs - Claim reported - Claim payment made - Claim file closed. **[1]**
Any other sensible comment award one mark to a max of four.

(b) Under a proportional reinsurance arrangement, the ceding company and the reinsurer share in all aspects of the risk. Most importantly, this means they share in the cost of claims for the risk.The cedant must pay a premium for this. The premium calculations will be based on the reinsurance treaty. **[1]**
Examples are:
Quota share reinsurance: Where the proportions are the same for all risks.
Surplus reinsurance: Where the proportions can vary from risk to risk. **[1]**
With non-proportional reinsurance the cedant pays a fixed premium to the reinsurer. The reinsurer will then pay the part of the claim that lies in a particular reinsurance layer. The layer itself will be defined by a lower limit called the retention limit and an upper limit.The upper limit can be infinite if the cover is unlimited. **[1]**
Examples are:
Individual Excess of Loss: The reinsurer will make payment when the amount of a claim exceeds a specified excess point. **[1]**
For example a reinsurer might pay the excess when the amount of a particular claim from a car insurance policy exceeds £50,000 but to an upper limit of £1m **[1]**
Stop Loss: With Stop Loss insurance the reinsurer will make a payment if the total claim amount for a specified group of policies exceeds a specified amount.The amount is usually expressed as a percentage of the premium. **[1]**
For example a reinsurer might agree to pay 85% of the excess when the total claim amount for all car insurance policies exceeds 100% of the total premium with no upper limit. **[1]**

Maximum of four marks available. Two marks must be for definition of proportional, non-proportional and two for the examples. One mark for either defining the specific type of reinsurance or giving an example of its use

---

End of Paper – An appendix of 2 pages follows.

Statistics – Common Distributions

Discrete Distributions

Distribution	Density	Range of Variates	Mean	Variance
Uniform	$\frac{1}{N}$	$N = 1,2,\ldots$ $x = 1,2,\ldots,N$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$
Bernoulli	$p^x(1-p)^{1-x}$	$0 \le p \le 1, x = 0,1$	p	$p(1-p)$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$0 \le p \le 1, n = 1,2,\ldots$ $x = 0,1,\ldots n$	np	$np(1-p)$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda > 0, x = 0,1,2,\ldots$	λ	λ
Geometric	$p(1-p)^x$	$0 < p \le 1, x = 0,1,2,\ldots$	$\frac{(1-p)}{p}$	$\frac{(1-p)}{p^2}$

Continuous Distributions

Distribution	Density	Range of Variates	Mean	Variance
Uniform	$\frac{1}{b-a}$	$-\infty < a < b < \infty$ $a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal $N(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp[\frac{-(x-\mu)^2}{2\sigma^2}]$	$-\infty < \mu < \infty$ $\sigma > 0, -\infty < x < \infty$	μ	σ^2
Lognormal (μ,σ^2)	$\frac{1}{x\sqrt{2\pi\sigma^2}}\exp[\frac{-(logx-\mu)^2}{2\sigma^2}]$	$-\infty < \mu < \infty$ $\sigma > 0, -\infty < x < \infty$	$e^{(\mu+\frac{1}{2}\sigma^2)}$	$e^{(2\mu+\sigma^2)}(e^{\sigma^2}-1)$
Exponential	$\lambda e^{-\lambda x}$	$\lambda > 0, x \ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma (α,λ)	$\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}$	$\lambda > 0, \alpha > 0, x > 0$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
Weibull (c,γ)	$c\gamma x^{\gamma-1}e^{-cx^\gamma}$	$c > 0, \gamma > 0, x > 0$	$c^{-\frac{1}{\gamma}}\Gamma(1+\gamma^{-1})$	$c^{-\frac{2}{\gamma}}[\Gamma(1+2\frac{1}{\gamma})$ $-\Gamma^2(1+\frac{1}{\gamma})]$
Pareto(α,λ)	$\frac{\alpha\lambda^\alpha}{(\lambda+x)^{\alpha+1}}$	$\alpha > 0, \lambda > 0, x > 0$	$\frac{\lambda}{(\alpha-1)}$	$\frac{\lambda^2\alpha}{(\alpha-1)^2(\alpha-2)}$
Burr(α,λ,γ)	$\frac{\alpha\gamma\lambda^\alpha x^{\gamma-1}}{(\lambda+x^\gamma)^{\alpha+1}}$	$\alpha > 0, \lambda > 0, \gamma > 0, x > 0$	Not required	Not required

Useful Formulae

EBCT Model 1

$$E[m(\theta)] = \bar{X}$$

$$E[s^2(\theta)] = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(n-1)}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2$$

$$var[m(\theta)] = \frac{1}{(N-1)}\sum_{i=1}^{N}(\bar{X}_i-\bar{X})^2 - \frac{1}{Nn}\sum_{i=1}^{N}\frac{1}{(n-1)}\sum_{j=1}^{n}(X_{ij}-\bar{X}_i)^2$$

EBCT Model 2

$$E[m(\theta)] = \bar{X}$$

$$E[s^2(\theta)] = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(n-1)}\sum_{j=1}^{n}P_{ij}(X_{ij}-\bar{X}_i)^2$$

$$var[m(\theta)] = \frac{1}{P^*}\left[\frac{1}{Nn-1}\sum_{i=1}^{N}\sum_{j=1}^{n}P_{ij}(X_{ij}-\bar{X})^2 - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{(n-1)}\sum_{j=1}^{n}P_{ij}(X_{ij}-\bar{X}_i)^2\right]$$

Intermediate calculations

$\sum_{j=1}^{n} P_{ij} = \bar{P}_i$ $\quad \sum_{i=1}^{N} \bar{P}_i = \bar{P}$ $\quad \frac{1}{(Nn-1)}\sum_{i=1}^{N}\bar{P}_i\left(1-\frac{\bar{P}_i}{\bar{P}}\right) = P^*$

$\sum_{j=1}^{n}\frac{P_{ij}X_{ij}}{\bar{P}_i} = \bar{X}_i$ $\quad \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{P_{ij}X_{ij}}{\bar{P}} = \bar{X}$

End of Appendix.