Stat 111 Presentation: Poisson and Chi Square

Mathematical Statistics 2, Spring 2025 Caroline Yao

We recall from Presentation 1 that if $X_1 \sim Pois(\lambda_1)$ is independent of $X_2 \sim Pois(\lambda_2)$, and $N = X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$, then $X_1 \mid N = n \sim Binom(n, p)$, for $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$.
We want to show that with the Poisson representation, the Pearson Chi-square statistic is the sum of squared standardized Poisson variables.
First recall that the Pearson Chi-squared test statistic is defined as

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i is observed data and E_i is the expected data. We know that for a Poisson distribution with λ, the mean and variance are both each to λ. So we can rewrite the above equation as

$$\chi^2 = \sum_i \frac{(O_i - \lambda)^2}{\lambda} = \sum_i \frac{(O_i - \lambda)^2}{(\sqrt{\lambda})^2} = \sum_i (\frac{O_i - \lambda}{\sqrt{\lambda}})^2$$

It is easy to recognize that $\frac{O_i - \lambda}{\sqrt{\lambda}}$ is the standardized Poisson variable. This may sound like it suggests that we have a chi-square distribution with degree of freedom 4, but this is not true because they are not independent from each other.

Now, let's suppose that we have a 2×2 table and the four counts are independent $Pois(\theta_{ij})$ variables, where $i = 1, 2$ and $j = 1, 2$. We want to show that conditioning on the row totals result in Binomial variables with probabilities $\phi_1 = \frac{\theta_{11}}{\theta_{11}+\theta_{12}}$ and $\phi_2 = \frac{\theta_{21}}{\theta_{21}+\theta_{22}}$.

	Column 1	Column 2	Row total
Row 1	$Pois(\theta_{11})$	$Pois(\theta_{12})$	T_1
Row 2	$Pois(\theta_{21})$	$Pois(\theta_{22})$	T_2

Table 1: 2 by 2 table where the four counts are independent $Pois(\theta_{ij})$ variables

Then denote $Pois(\theta_{ij})$ as X_{ij}. By part a, we know that $T_1 \sim Pois(\theta_{11} + \theta_{12})$ and $X_{11} \mid T_1 = t \sim Bin(t, \phi_1)$ with $\phi_1 = \frac{\theta_{11}}{\theta_{11}+\theta_{12}}$. Similarly, $T_2 \sim Pois(\theta_{21} + \theta_{22})$ and $X_{21} \mid T_2 = t \sim Bin(t, \phi_2)$ with $\phi_2 = \frac{\theta_{21}}{\theta_{21}+\theta_{22}}$.

We know that Jeffrey's non-informative prior for a Poisson rate θ is $p(\theta) \propto \theta^{-1/2} I(\theta > 0)$. Let's assume independent priors for the four Poisson rates, we want to show that the joint posterior density for the θ_{ij}'s is that of four independent Gamma$(X_{ij} + 1/2, 1)$ random variables, and for ϕ_1 and ϕ_2 it is that of two independent $Beta(X_{i1} + 1/2, X_{i2} + 1/2)$ variables.
We know that posterior $\propto L(\theta_{ij}) p_\theta(\theta)$ and $L(\theta_{ij}) = \frac{\theta_{ij}^{X_{ij}}}{X_{ij}!} e^{-\theta_{ij}}$ and we are given that $p(\theta_{ij}) \propto$

$\theta_{ij}^{-1/2} I(\theta_{ij} > 0)$.

$$\begin{aligned} p_{\theta_{ij}|X_{ij}} &= \frac{\theta_{ij}^{X_{ij}}}{X_{ij}!} e^{-\theta_{ij}} \cdot \theta_{ij}^{-1/2} I(\theta_{ij} > 0) \\ &\propto \theta_{ij}^{X_{ij}-1/2} e^{-\theta_{ij}} \\ &\sim Gamma(X_{ij} + 1/2, 1) \end{aligned}$$

As for ϕ_1 and ϕ_2, recall that we found $\phi_1 = \frac{\theta_{11}}{\theta_{11}+\theta_{12}}$ and $\phi_2 = \frac{\theta_{21}}{\theta_{21}+\theta_{22}}$. From Week 2 presentation problem 4, we know that if $V_1 \sim Gamma(a, \lambda)$ and $V_2 \sim Gamma(b, \lambda)$ independent from each other, then $\frac{V_1}{V_1+V_2} \sim Beta(a, b)$. So here, we immediately have $\phi_1 \sim Beta(X_{11} + 1/2, X_{12} + 1/2)$ and $\phi_2 \sim Beta(X_{21} + 1/2, X_{22} + 1/2)$.
We see that this agrees with assuming independent $Beta(1/2, 1/2)$ prior densities for ϕ_1 and ϕ_2.
We see that $X_{11} \mid X_{11} + X_{12} \sim Bin(X_{11} + X_{12}, \phi_1)$. So we find the likelihood to be $L(\phi_1) = p(X_{11} \mid X_{11} + X_{12}, \phi_1) \propto \phi_1^{X_{11}}(1 - \phi_1)^{X_{12}}$.

$$\begin{aligned} p(\phi_1 \mid X_{11}, X_{12}) &\propto \phi_1^{X_{11}}(1-\phi_1)^{X_{12}} \cdot \phi_1^{-1/2} \cdot (1-\phi_1)^{-1/2} \\ &\propto \phi_1^{X_{11}-1/2}(1-\phi_1)^{X_{12}-1/2} \\ &\sim Beta(X_{11} + 1/2, X_{12} + 1/2) \end{aligned}$$

We leave the case for ϕ_2 to the readers.

Let's find a Bayes posterior 95% interval for $\phi_1 - \phi_2$ for the coffee data and compare to the large-sample CI.
According to the data given, $\phi_1 \sim Beta(34.5, 80-34+1/2) = Beta(34.5, 46.5)$ and $\phi_2 \sim Beta(24.5, 40-24+1/2) = Beta(24.5, 16.5)$.
We use R to simulate data to find the posterior. We use the following code given by Phil:

```
x11 = 34; x12 = 80-34; x21 = 24; x22 = 40-24
phi1 = rbeta(100000, x11+0.5, x12 + 0.5); phi2 = rbeta(100000, x21+0.5, x22 + 0.5)
quantile(phi2-phi1, c(0.025, 0.975)
```

We should get $(-0.015, 0.351)$ as the final answer.
As for large sample, we find the proportion to be $p_1 = \frac{34}{80}$ and $p_2 = \frac{24}{40}$. We find $SE = \sqrt{\frac{0.425\cdot(1-0.425)}{80} + \frac{0.6\cdot(1-0.6)}{40}} = 0.095$. With $z^* = 1.96$, we get $(0.6 - 0.425) \pm 1.96 \cdot 0.095 = (-0.0112, 0.3612)$. We see that these intervals are consistent with each other.

Finally, let's carry out a chi-square test on the data provided.

	0 cups	1-4 cups	5 or more	Total
Fr	29 [24.375]	11 [11.625]	5 [9]	45
So	21 [18.96]	9 [9.042]	5 [7]	35
Jr	10 [13.54]	7 [6.458]	8 [5]	25
Sr	5 [8.125]	4 [3.875]	6 [3]	15
	65	31	24	120

Table 2: Data provided with expected value calculated in brackets

We first calculated the expected value as row total divide by total and multiply by column total. Next, we apply the Chi-square formula.

$$\chi^2 = \sum_{i,j} \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(29-24.375)^2}{24.375} + \frac{(11-11.625)^2}{11.625} + \frac{(5-9)^2}{9} + \frac{(21-18.96)^2}{18.96} + \frac{(9-9.042)^2}{9.042} + \frac{(5-7)^2}{7} + \frac{(10-13.54)^2}{13.54} + \frac{(7-6.458)^2}{6.458} + \frac{(8-5)^2}{5} + \frac{(5-8.125)^2}{8.125} + \frac{(4-3.875)^2}{3.875} + \frac{(6-3)^2}{3} = 10.458$$

We know that the degree of freedom is $(row-1)\times(column-1)$. So here it is $(4-1)\times(3-1)=6$. This has a p-value of 0.1066, which is not significant at $p<0.05$.
We can try to tell what contributed most to the plot by finding out the signed square roots of the chi-square contributions. We see that SR with 5 or more, JR with 5 or more, FR with 5 or mote and SR with 0 cups relatively contributed more to the Chi-square statistics. Also notice that the middle column (1-4 cups) are relatively low and they have approximately the same height across different categories on the plot.

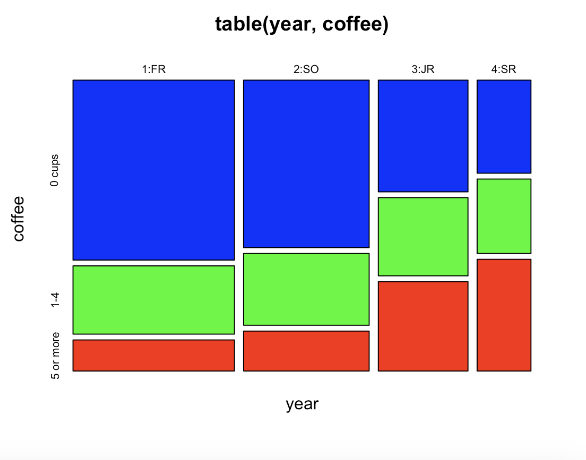

Figure 1: Mosaic Plot

year \ coffee	0 cups	1-4	5 or more
1:FR	0.93678391	-0.18330889	-1.33333333
2:SO	0.46890489	-0.01385685	-0.75592895
3:JR	-0.96243548	0.21314340	1.34164079
4:SR	-1.09632252	0.06350006	1.73205081

Figure 2: Signed square roots of the chi-square contributions