# Math 362: Mathematical Statistics II 

Le Chen<br>le.chen@emory.edu<br>Emory University Atlanta, GA

Last updated on April 13, 2021

2021 Spring

# Chapter 10. Goodness-of-fit Tests 

§ 10.1 Introduction
§ 10.2 The Multinomial Distribution
§ 10.3 Goodness-of-Fit Tests: All Parameters Known
§ 10.4 Goodness-of-Fit Tests: Parameters Unknown
§ 10.5 Contingency Tables

## Plan

§ 10.1 Introduction
§ 10.2 The Multinomial Distribution
§ 10.3 Goodness-of-Fit Tests: All Parameters Known
§ 10.4 Goodness-of-Fit Tests: Parameters Unknown
§ 10.5 Contingency Tables

# Chapter 10. Goodness-of-fit Tests 

§ 10.1 Introduction
§ 10.2 The Multinomial Distribution
§ 10.3 Goodness-of-Fit Tests: All Parameters Known
§ 10.4 Goodness-of-Fit Tests: Parameters Unknown
§ 10.5 Contingency Tables

| $p_{i}$ are known | $p_{i}$ are unknown |
| :---: | :---: |
| $D=\sum_{i=1}^{t} \frac{\left(X_{i}-n p_{i}\right)^{2}}{n \rho_{i}}$ | $D_{1}=\sum_{i=1}^{t} \frac{\left(X_{i}-n \hat{p}_{i}\right)^{2}}{n \hat{p}_{i}}$ |
| $\chi^{2}$ with f.d. $t-1$ | $\chi^{2}$ with f.d. $t-1-s$ |
| $d=\sum_{i=1}^{t} \frac{\left(k_{i}-n p_{i}\right)^{2}}{n p_{i 0}}$ | $d_{1}=\sum_{i=1}^{t} \frac{\left(k_{i}-n \hat{p}_{i 0}\right)^{2}}{n \hat{p}_{i 0}}$ |
| $n p_{i 0} \geq 5$ | $n \hat{p}_{i 0} \geq 5$ |
| $d>\chi_{1-\alpha, t-1}^{2}$ | $d_{1}>\chi_{1-\alpha, t-1-s}^{2}$ |

$\dagger s$ is the number of unknown parameters.

$$
\mathrm{df}=\underline{\text { number of classes }}-1-\underline{\text { number of unknown parameters }}
$$

E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let $X_{i}$ be the number of hits for the ith student.

> People believe that $X_{i}$ should following binomial $(4, p)$, that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors. Find the NTE for $p$. Use the data to make a conclusion.

Sol. 1) $H_{0}: X_{i} \sim \operatorname{binomal}(4, p)$.
2) Under $H_{0}$, the MLE for $p$ is $p_{e}=\ldots=0.251$
E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let $X_{i}$ be the number of hits for the ith student.

| Number of Hits, $i$ |
| ---: |
| $r_{i}^{\prime s}\left\{\begin{array}{lc}0 & \text { Obs. Freq., } k_{i} \\ 1 & 1280 \\ 2 & 1717 \\ 3 & 915 \\ 4 & 167 \\ \end{array}\right)$ |

People believe that $X_{i}$ should following binomial $(4, p)$, that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors.

Find the MLE for $p$. Use the data to make a conclusion

Sol. 1) $H_{0}: X_{i} \sim \operatorname{binomal}(4, p)$.
2) Under Ho. the MTE for $\rho$ is $\rho_{e}=\ldots=0.251$
E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let $X_{i}$ be the number of hits for the ith student.

| Number of Hits, $i$ |
| :---: |
| $r_{i}^{\prime s}$ Obs. Freq., $k_{i}$ |
| $\left(\begin{array}{lc}0 & 1280 \\ 1 & 1717 \\ 2 & 915 \\ 3 & 167 \\ 4 & 17\end{array}\right.$ |

People believe that $X_{i}$ should following binomial $(4, p)$, that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors.

Find the MLE for $p$. Use the data to make a conclusion.

Sol. 1) $H_{0}: X_{i} \sim \operatorname{binomal}(4, p)$.
2) Tnder Ho. the MTE for $\rho$ is $\rho_{e}=\ldots=0.251$
E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let $X_{i}$ be the number of hits for the ith student.

| Number of Hits, $i$ |
| :---: |
| $r_{i}^{\prime s}$ Obs. Freq., $k_{i}$ |
| $\left(\begin{array}{lc}0 & 1280 \\ 1 & 1717 \\ 2 & 915 \\ 3 & 167 \\ 4 & 17\end{array}\right.$ |

People believe that $X_{i}$ should following binomial $(4, p)$, that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors.

Find the MLE for $p$. Use the data to make a conclusion.

Sol. 1) $H_{0}: X_{i} \sim \operatorname{binomal}(4, p)$.
2) Under $H_{0}$, the MLE for $p$ is $p_{e}=$
E.g. 1 Binomial data: 4096 students, each shots basketball 4 times. Let $X_{i}$ be the number of hits for the ith student.

| Number of Hits, $i$ |
| :---: |
| $r_{i}^{\prime s}$ Obs. Freq., $k_{i}$ |
| $\left(\begin{array}{ll}0 & 1280 \\ 1 & 1717 \\ 2 & 915 \\ 3 & 167 \\ 4 & 17\end{array}\right.$ |

People believe that $X_{i}$ should following binomial $(4, p)$, that is, shotting basketball should be something like trying to get red chocolate beans from a jar of beans of two colors.

Find the MLE for $p$. Use the data to make a conclusion.

Sol. 1) $H_{0}: X_{i} \sim \operatorname{binomal}(4, p)$.
2) Under $H_{0}$, the MLE for $p$ is $p_{e}=\ldots=0.251$
3) Compute the expected frequenies:

$$
\Longrightarrow \quad d_{1}=\cdots=6.401 .
$$

4) Critical region: $\left(\chi_{.95,5-1-1}^{2},+\infty\right)=(7.815,+\infty)$
5) Conclusion: Fail to reject.
6) Alternatively, $P$-value $=\mathbb{P}\left(\chi_{3}^{2} \geq 6.401\right)=0.094, \ldots$ discuss...
7) Compute the expected frequenies:

| Table IO.4.I |  |  |
| :---: | :---: | :---: |
| Number of Hits, $i$ | Obs. Freq., $k_{i}$ | Estimated Exp. Freq., $n \hat{p}_{i_{o}}$ |
| $r_{i}^{\prime} s$ | 1280 | 1289.1 |
| 1 | 1717 | 1728.0 |
| 2 | 915 | 868.6 |
| 3 | 167 | 194.0 |
| 4 | 17 | 16.3 |
|  |  | $d_{1}=\cdots=6.401$. |

4) Critical region: $\left(\chi_{.95,5-1-1}^{2},+\infty\right)=(7.815,+\infty)$
5) Compute the expected frequenies:

| Table 10.4.I |  |  |
| :--- | :---: | :---: |
| Number of Hits, $i$ | Obs. Freq., $k_{i}$ | Estimated Exp. Freq., $n \hat{p}_{i_{o}}$ |
| $r_{i}^{\prime} s$ | 1280 | 1289.1 |
| 1 | 1717 | 1728.0 |
| 2 | 915 | 868.6 |
| 3 | 167 | 194.0 |
| 4 | 17 | 16.3 |
|  |  | $d_{1}=\cdots=6.401$. |

4) Critical region: $\left(\chi_{.95,5-1-1}^{2},+\infty\right)=(7.815,+\infty)$
5) Conclusion: Fail to reject.
6) Compute the expected frequenies:

| Table 10.4.I |  |  |
| :--- | :---: | :---: |
| Number of Hits, $i$ | Obs. Freq., $k_{i}$ | Estimated Exp. Freq., $n \hat{p}_{i_{o}}$ |
| $r_{i}^{\prime} s$ | 1280 | 1289.1 |
| 1 | 1717 | 1728.0 |
| 2 | 915 | 868.6 |
| 3 | 167 | 194.0 |
| 4 | 17 | 16.3 |
|  |  | $d_{1}=\cdots=6.401$. |

4) Critical region: $\left(\chi_{.95,5-1-1}^{2},+\infty\right)=(7.815,+\infty)$
5) Conclusion: Fail to reject.
6) Alternatively, $P$-value $=\mathbb{P}\left(\chi_{3}^{2} \geq 6.401\right)=0.094, \ldots$ discuss...
E.g. 2 Does the number of death per day follow the Poisson distribution?
E.g. 2 Does the number of death per day follow the Poisson distribution?

| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ |
| :---: | :---: |
| 0 | 162 |
| 1 | 267 |
| 2 | 271 |
| 3 | 185 |
| 4 | 111 |
| 5 | 61 |
| 6 | 27 |
| 7 | 8 |
| 8 | 3 |
| 9 | 1 |
| $10+$ | 0 |
|  | 1096 |

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.
3) The MLE for $\lambda$ is: $\lambda_{e}=\cdots=2.157$.

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.
3) The MLE for $\lambda$ is: $\lambda_{e}=\cdots=2.157$.
4) Compute the expected frequencies:

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.
3) The MLE for $\lambda$ is: $\lambda_{e}=\cdots=2.157$.
4) Compute the expected frequencies:

## Table 10.4.2

| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{o}}$ |
| :---: | :---: | :---: |
| 0 | 162 | 126.8 |
| 1 | 267 | 273.5 |
| 2 | 271 | 294.9 |
| 3 | 185 | 212.1 |
| 4 | 111 | 114.3 |
| 5 | 61 | 49.3 |
| 6 | 27 | 17.8 |
| 7 | 8 | 5.5 |
| 8 | 3 | 1.4 |
| 9 | 1 | 0.3 |
| $10+$ | 0 | 0.1 |
|  | 1096 | 1096 |


| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{o}}$ |
| :---: | :---: | :---: |
| $r_{1}, r_{2}, \ldots, r_{8}\left\{\begin{array}{l}0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7+\end{array}\right.$ | 162 | 126.8 |
|  | 267 | 273.5 |
|  | 271 | 294.9 |
|  | 185 | 212.1 |
|  | 111 | 114.3 |
|  | 61 | 49.3 |
|  | 27 | 17.8 |
|  | 12 | 7.3 |
|  | 1096 | 1096 |

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.
3) The MLE for $\lambda$ is: $\lambda_{e}=\cdots=2.157$.
4) Compute the expected frequencies:

## Table 10.4.2

| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{o}}$ |
| :---: | :---: | :---: |
| 0 | 162 | 126.8 |
| 1 | 267 | 273.5 |
| 2 | 271 | 294.9 |
| 3 | 185 | 212.1 |
| 4 | 111 | 114.3 |
| 5 | 61 | 49.3 |
| 6 | 27 | 17.8 |
| 7 | 8 | 5.5 |
| 8 | 3 | 1.4 |
| 9 | 1 | 0.3 |
| $10+$ | 0 | 0.1 |
|  | 1096 | 1096 |


| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{o}}$ |
| :---: | :---: | :---: |
| $r_{1}, r_{2}, \ldots, r_{8}\left\{\begin{array}{l}0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7+\end{array}\right.$ | 162 | 126.8 |
|  | 267 | 273.5 |
|  | 271 | 294.9 |
|  | 185 | 212.1 |
|  | 111 | 114.3 |
|  | 61 | 49.3 |
|  | 27 | 17.8 |
|  | 12 | 7.3 |
|  | 1096 | 1096 |

$$
\Longrightarrow \quad d_{1}=\cdots=25.98 .
$$

Sol. 1) Let $X_{i}$ be the number of death in $i$ th day, $1 \leq i \leq 1096$.
2) $H_{0}: X_{i}$ follow Poisson $(\lambda)$.
3) The MLE for $\lambda$ is: $\lambda_{e}=\cdots=2.157$.
4) Compute the expected frequencies:

| Table 10.4.2 |  |  |
| :---: | :---: | :---: |
| Number of Deaths, $i$ | Obs. Freq., $\boldsymbol{k}_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{i}}$ |
| 0 | 162 | 126.8 |
| 1 | 267 | 273.5 |
| 2 | 271 | 294.9 |
| 3 | 185 | 212.1 |
| 4 | 111 | 114.3 |
| 5 | 61 | 49.3 |
| 6 | 27 | 17.8 |
| 7 | 8 | 5.5 |
| 8 | 3 | 1.4 |
| 9 | 1 | 0.3 |
| $10+$ | $\underline{1096}$ | $\underline{0.1}$ |
|  |  | 1096 |


| Number of Deaths, $i$ | Obs. Freq., $k_{i}$ | Est. Exp. Freq., $n \hat{p}_{i_{e}}$ |
| :---: | :---: | :---: |
| $r_{1}, r_{2}, \ldots, r_{8}\left\{\begin{array}{l}0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7+\end{array}\right.$ | 162 | 126.8 |
|  | 267 | 273.5 |
|  | 271 | 294.9 |
|  | 185 | 212.1 |
|  | 111 | 114.3 |
|  | 61 | 49.3 |
|  | 27 | 17.8 |
|  | 12 | 7.3 |
|  | 1096 | 1096 |

$$
\Longrightarrow \quad d_{1}=\cdots=25.98 .
$$

5) $P$-value $=\mathbb{P}\left(\chi_{1,8-1-1}^{2} \geq 25.98\right)=0.00022$. Reject!
