浅草夜見世

Statistics
math 414

Lecture 11

Vivek Tewary

Fitting of Probability Distribution

The emission of alpha particles from radioactive sources is often modeled by Poisson distribution.

n	Observed	Expected
0–2	18	12.2
3	28	27.0
4	56	56.5
5	105	94.9
6	126	132.7
7	146	159.1
8	164	166.9
9	161	155.6
10	123	130.6
11	101	99.7
12	74	69.7
13	53	45.0
14	23	27.0
15	15	15.1
16	9	7.9
17+	5	7.1
	1207	1207

The experimenters record times between successive emissions. In the table, n is the number of emissions observed in 1207 intervals.

For examples, there were 3 emissions in 28 intervals of 10 seconds. In 56 of the intervals, there were 4 emissions.

In fitting the model, each interval is seen as an independent realization of a Poisson random variable with p.m.f.

$$\pi_k = P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

In fitting the model, each interval is seen as an independent realization of a Poisson random variable with p.m.f.

$$\pi_k = P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

In this fitting, we want to find λ. Since the average count in each of the interval is 8.392, we take this as an "estimate" for λ. We denote it by $\hat{\lambda}$. Clearly the estimate itself is a random variable.

Under this model the probability that an observation falls in the first row (0, 1 or 2 counts) is

$$\pi_0 + \pi_1 + \pi_2 = p_1 .$$

3 counts is π_3 and so on. In all, the model follows a multinomial probability distribution with $n = 1207$ and probabilities $p_1, p_2, \ldots, p_{16}$, calculated as above using the Poisson distribution. The third column tabulates these numbers.

Methods of Estimation

① Method of moments.

Methods of Estimation

① Method of moments.

The k^{th} moment of a distribution is defined as $\mu_k = E(X^k)$.

The k^{th} sample moment for iid random variables $X_1, X_2, \dots, X_n$ is defined as

$$\boxed{\hat{\mu}_k = \frac{\sum_{i=1}^{n} X_i^k}{n}}$$

Methods of Estimation

① Method of moments.

The k^{th} moment of a distribution is defined as $\mu_k = E(X^k)$.

The k^{th} sample moment for iid random variables $X_1, X_2, \ldots, X_n$ is defined as

$$\boxed{\hat{\mu}_k = \frac{\sum_{i=1}^{n} X_i^k}{n}}$$

an estimate for μ_k.

In the method of moments, we estimate parameters associated with a random variable by expersing those parameters in terms of moments and then substituting the sample moments in their place.

For example, for Poisson distribution, the parameter of interest is λ which is also $E[X]$ and hence, the first sample moment $\hat{\mu}_1$ is an estimate.

Example.

Consider the following data

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Let us fit Poisson distribution to this data

$$\hat{\lambda} = \frac{\sum x_i}{n} = 24.9.$$

$\hat{\lambda}$ is a random variable which is approximately normal for large n. Its standard deviation is approximately $\sqrt{\frac{\hat{\lambda}}{n}}$ under the Poisson approximation.

Example.

Consider the following data

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Whether Poisson is a "good fit" will be addressed later.

Let us fit Poisson distribution to this data

$$\hat{\lambda} = \frac{\sum x_i}{n} = 24.9.$$

$\hat{\lambda}$ is a random variable which is approximately normal for large n. Its standard deviation is approximately $\sqrt{\frac{\hat{\lambda}}{n}}$ under the Poisson approximation.

Example 2 In the case of the normal distribution $N(\mu, \sigma^2)$

$$\mu_1 = E(X) = \mu$$

$$\mu_2 = E(X^2) = \mu^2 + \sigma^2$$

$$\therefore \quad \mu = \mu_1$$

$$\sigma^2 = \mu_2 - \mu_1^2$$

Therefore $\boxed{\hat{\mu} = \overline{X}}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

Example 2

In the case of the normal distribution $N(\mu, \sigma^2)$

$$\mu_1 = E(X) = \mu$$

$$\mu_2 = E(X^2) = \mu^2 + \sigma^2$$

Sampling distribution of $\hat{\mu}$ is $N(\mu, \sigma^2/n)$

$$\therefore \quad \mu = \mu_1$$

$$\sigma^2 = \mu_2 - \mu_1^2$$

Therefore $\hat{\mu} = \overline{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Sampling distribution of $\frac{n\hat{\sigma}^2}{\sigma^2}$ is χ^2_{n-1}.

Example 3 Even when we know the probability distribution of a random variable explicitly, it is often difficult to find the sampling distributions for its parameters. In such cases, we can use simulation. Consider the Gamma (α, β) distribution with parameters $\alpha > 0$ and $\beta > 0$. Its p.d.f is

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} & , 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

$\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$

Example 3 Even when we know the probability distribution of a random variable explicitly, it is often difficult to find the sampling distributions for its parameters. In such cases, we can use simulation. Consider the Gamma (α, β) distribution with parameters $\alpha > 0$ and $\beta > 0$. Its p.d.f is

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

$$\mu = \alpha\beta \quad \text{and} \quad \sigma^2 = \alpha\beta^2$$

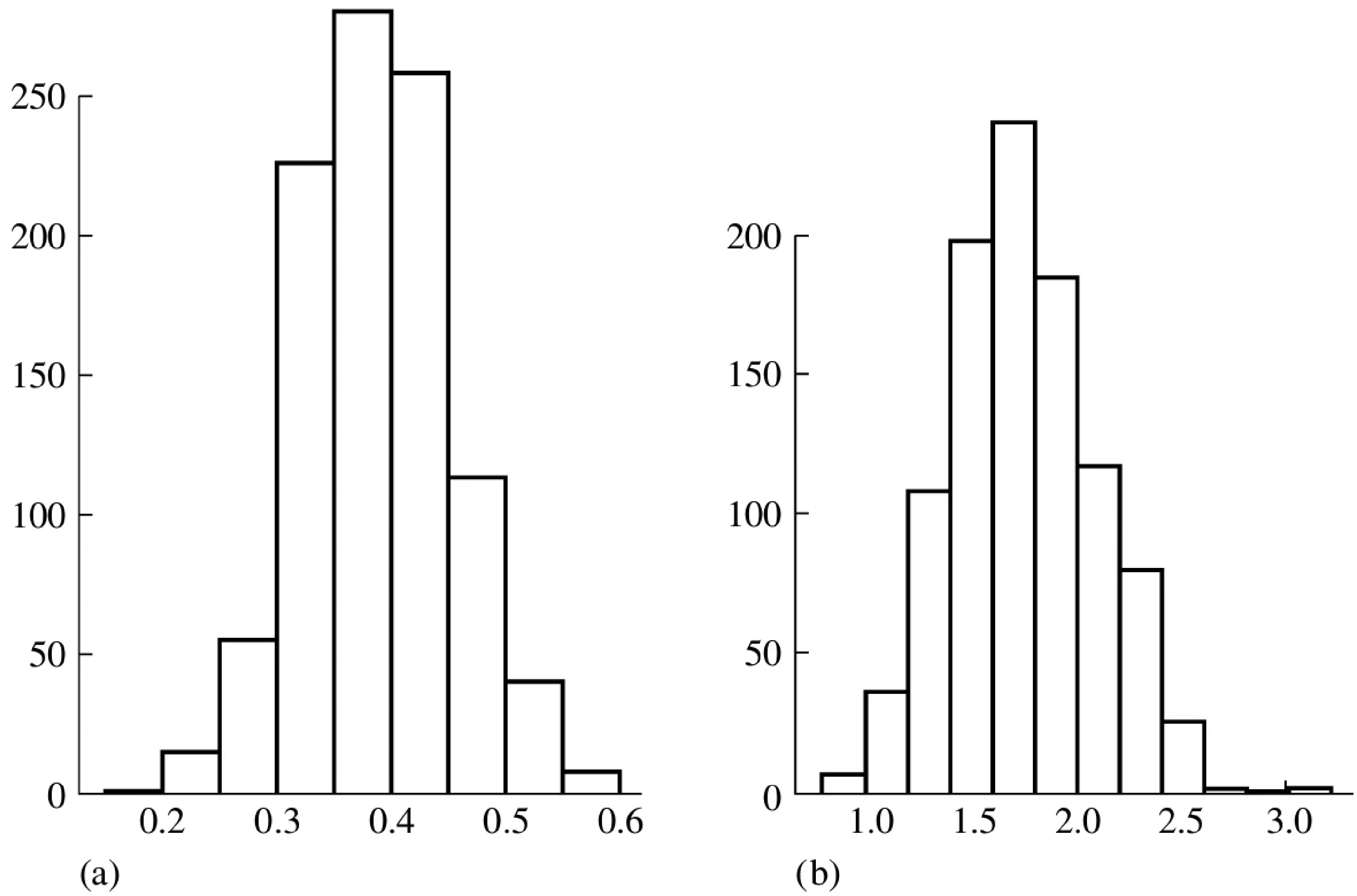
By the method of moments.

$$\mu_1 = \alpha\beta,\ \mu_2 = \alpha^2\beta^2 + \alpha\beta^2 = \mu_1^2 + \mu_1\beta \Rightarrow \beta = \frac{\mu_2 - \mu_1^2}{\mu_1},\ \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

Let α_0 and β_0 be the true values of α and β.

To produce the simulations, we would need to know the true values α_0 and β_0.

- Suppose that we draw 1000 samples of size $n = 227$ with $\alpha = .375$ and $\beta = 0.597$
- Using the method of moments, we find $\hat{\alpha}, \hat{\beta}$ and plot their histograms.
- At this point, we could use the normal approx: for their sampling distributions.
- We could use the standard deviations of these estimates as a measure of their variability.


250
200
150
100
50
0
0.2
0.3
0.4
0.5
0.6
(a)
200
150
100
50
0
1.0
1.5
2.0
2.5
3.0
(b)

Consistency

An estimate $\hat{\theta}$ is said to be a consistent estimate of a parameter θ if $\hat{\theta}$ approaches θ as the sample size approaches infinity. More precisely,

Let $\hat{\theta}_n$ be an estimate of a parameter θ based on a sample of size n. Then $\hat{\theta}_n$ is consistent in probability if $\hat{\theta}_n$ converges in probability to θ as $n \longrightarrow \infty$, i.e., for any $\varepsilon > 0$,

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty.$$

The sample moments are consistent estimators of the true moments. This is a consequence of Chebyshev's inequality.

Specifically

$$E(\overline{X_n^k}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^k) = \mu_k$$

and $$Var(\overline{X_n^k}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i^k) = \frac{Var(X^k)}{n}$$

$$\therefore \quad P(|\overline{X_n^k} - \mu_k| > \varepsilon) \leq \frac{Var(X^k)}{n\varepsilon^2} \longrightarrow 0 \text{ as}$$

$$n \longrightarrow \infty.$$

The sample moments are consistent estimators of the true moments. This is a consequence of Chebyshev's inequality (or law of large numbers)

<u>Specifically</u>

$$E(\overline{X_n^k}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \mu_k$$

and $$Var(\overline{X_n^k}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i^k) = \frac{Var(X^k)}{n}$$

$$\therefore \quad P(|\overline{X_n^k} - \mu_k| > \varepsilon) \leq \frac{Var(X^k)}{n\varepsilon^2} \longrightarrow 0 \text{ as}$$

$$n \to \infty.$$

It is not difficult to see that "nice" functions of sample moments are also consistent estimators.

It is not difficult to see that "nice" functions of sample moments are also consistent estimators.

Therefore if λ_0 is the true parameter of a Poisson distribution X. Then $\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda_0}{n}}$.
We can approximate this by

$$S_{\hat{\theta}} = \sqrt{\frac{\hat{\lambda}}{n}}$$

because by the previous argument (involving law of large numbers)

$$\hat{\lambda} \longrightarrow \lambda_0 \quad \text{as} \quad n \longrightarrow \infty.$$

It is not difficult to see that "nice" functions of sample moments are also consistent estimators.

Therefore if λ_0 is the true parameter of a Poisson distribution X. Then $\sigma_{\hat{\lambda}} = \sqrt{\frac{\lambda_0}{n}}$.

We can approximate this by

$$S_{\hat{\theta}} = \sqrt{\frac{\hat{\lambda}}{n}}$$

because by the previous argument (involving law of large numbers)

$$\hat{\lambda} \longrightarrow \lambda_0 \quad \text{as} \quad n \longrightarrow \infty.$$

This shows that $S_{\hat{\theta}}$ is a useful measure of variability of $\hat{\lambda}$.

Method of maximum likelihood

Method of maximum likelihood

Recall that if $X_1, X_2, \ldots, X_n$ are random variables with joint density $f(x_1, x_2, \ldots, x_n; \theta)$, the likelihood of θ as a function of observed values $x_1, x_2, \ldots, x_n$ is defined as

$$\boxed{L(\theta) = f(x_1, x_2, \ldots, x_n ; \theta)}$$

The maximum likelihood estimate (mle) of θ is that value of θ that maximizes the likelihood function.

If the X_i are i.i.d. then

$$\boxed{L(\theta) = \prod_{i=1}^{n} f(x_i | \theta)}$$

Of course, we often maximize the log-likelihood function

$$\boxed{\ell(\theta) = \sum_{i=1}^{n} \log f(x_i | \theta)}$$

Earlier we found mle for <u>normal</u> distribution and <u>Binomial</u> distribution.

In particular, we found simple expressions for the mle. This is not always the case.

<u>Example</u> <u>Gamma distribution</u>

$$f(x\mid\alpha,\beta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty.$$

Earlier we found mle for normal distribution and Binomial distribution.

In particular, we found simple expressions for the mle. This is not always the case.

Example Gamma distribution

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\, \beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty.$$

Therefore, the log likelihood function is

$$l(\alpha, \beta) = \sum_{i=1}^{n} (\alpha - 1) \log x_i - \frac{1}{\beta} x_i - \log \Gamma(\alpha) - \alpha \log \beta$$

Earlier we found mle for normal distribution and Binomial distribution.

In particular, we found simple expressions for the mle. This is not always the case.

Example Gamma distribution

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty.$$

Therefore, the log likelihood function is

$$l(\alpha,\beta) = \sum_{i=1}^{n} (\alpha-1)\log x_i - \frac{1}{\beta} x_i - \log \Gamma(\alpha) - \alpha \log \beta$$

$$= (\alpha-1)\sum_{i=1}^{n} \log x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i - n \log\Gamma(\alpha) - n\alpha\log\beta$$

Example: Gamma distribution

$$f(x \mid \alpha, \beta) = \frac{1}{\Gamma(\alpha)\,\beta^{\alpha}}\, x^{\alpha-1} e^{-x/\beta}, \quad 0 \le x < \infty$$

Therefore, the log likelihood function is

$$l(\alpha,\beta) = \sum_{i=1}^{n} (\alpha-1)\log x_i - \frac{1}{\beta} x_i - \log\Gamma(\alpha) - \alpha\log\beta$$

$$= (\alpha-1)\sum_{i=1}^{n}\log x_i - \frac{1}{\beta}\sum_{i=1}^{n} x_i - n\log\Gamma(\alpha) - n\alpha\log\beta$$

The partial derivatives are

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n}\log x_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\log\beta$$

$$\frac{\partial l}{\partial \beta} = \frac{\sum_{i=1}^{n} x_i}{\beta^2} - \frac{n\alpha}{\beta}.$$

Example Gamma distribution

The partial derivatives are

$$\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{n} \log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n \log \beta$$

$$\frac{\partial l}{\partial \beta} = \frac{\sum_{i=1}^{n} X_i}{\beta^2} - \frac{n\alpha}{\beta}.$$

Setting the second eq. to 0, we get

$\hat{\beta} = \frac{\overline{X}}{\hat{\alpha}}$ where $\hat{\alpha}$ is to be found from the first expression set to zero.

$$\sum_{i=1}^{n} \log X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} - n \log \bar{X} + n \log \hat{\alpha} = 0$$

This cannot be solved explicitly. Must employ numerical techniques.

The iterative numerical techniques require an initial guess. In this example, this initial guess could be supplied by the method of moments. $(\hat{\alpha} = 0.441, \ \hat{\beta} = 0.510)$

Once again the question of the sampling distribution of the mle arises. In this case, the sampling distribution must be found by simulation which requires the knowledge of the true values of α and β. We can use the mle in their place.

Maximum likelihood Estimates for Multinomial Cell Probabilities

Suppose that $X_1, X_2, \ldots, X_m$, the counts in cells $1, \ldots, m$ follow a multinomial distribution with a total count of n and cell probabilities $p_1, p_2, \ldots, p_m$. We want to estimate p_i.

The joint p.m.f is

$$\boxed{f(x_1, x_2, \ldots, x_m \mid p_1, p_2, \ldots, p_m) = \frac{n!}{\prod_{i=1}^{m} x_i!} \prod_{i=1}^{m} p_i^{x_i}.}$$

The marginal distribution of each X_i is binomial (n, p_i). This random sample is not i.i.d. We can still use the m.l.e. method.

The marginal distribution of each X_i is binomial (n, p_i). This random sample is not i.i.d. We can still use the m.l.e. method.

The log likelihood function is

$$\boxed{l(p_1, p_2, \ldots, p_m) = \log n! - \sum_{i=1}^{n} \log x_i! + \sum_{i=1}^{n} x_i \log p_i}$$

The marginal distribution of each X_i is binomial (n, p_i). This random sample is not i.i.d. We can still use the m.l.e. method.

The log likelihood function is

$$\boxed{l(p_1, p_2, \ldots, p_m) = \log n! - \sum_{i=1}^{n} \log x_i! + \sum_{i=1}^{n} x_i \log p_i}$$

However, we have the constraint

$$\sum_{i=1}^{n} p_i = 1.$$

The marginal distribution of each X_i is binomial (n, p_i). This random sample is not i.i.d. We can still use the m.l.e. method.

The log likelihood function is

$$\boxed{l(p_1, p_2, \ldots, p_m) = \log n! - \sum_{i=1}^{m} \log x_i! + \sum_{i=1}^{m} x_i \log p_i}$$

However, we have the constraint

$$\sum_{i=1}^{m} p_i = 1.$$

We must use the Lagrange multiplier method.

Define

$$L(p_1, p_2, \dots, p_m; \lambda) = \log n! - \sum_{i=1}^{m} \log x_i! + \sum_{i=1}^{m} x_i \log p_i + \lambda\left(\sum_{i=1}^{m} p_i - 1\right)$$

Define

$$L(p_1, p_2, \ldots, p_m; \lambda) = \log n! - \sum_{i=1}^{m} \log x_i! + \sum_{i=1}^{m} x_i \log p_i + \lambda\left(\sum_{i=1}^{m} p_i - 1\right)$$

Setting the partial derivatives equal to zero yields:

$$\hat{p}_j = -\frac{x_j}{\lambda}.$$

Sum over j to get

$$1 = \frac{-n}{\lambda} \quad \text{or} \quad \boxed{\lambda = -n}$$

$$\therefore \boxed{\hat{p}_j = x_j/n}$$

Define

$$L(p_1, p_2, \ldots, p_m; \lambda) = \log n! - \sum_{i=1}^{m} \log x_i! + \sum_{i=1}^{m} x_i \log p_i + \lambda\left(\sum_{i=1}^{m} p_i - 1\right)$$

Setting the partial derivatives equal to zero yields:

$$\hat{p}_j = -\frac{x_j}{\lambda}.$$

Sum over j to get

$$1 = \frac{-n}{\lambda} \quad \text{or} \quad \boxed{\lambda = -n}$$

$$\therefore \boxed{\hat{p}_j = x_j/n}$$

This is "obvious".