ST2132 Parameter Estimation I

Fitting distribution, Method of moments

Semester 1 2022/2023

Parameter estimation

- What RV's might have generated the data? The statistician faces this question very often. The goal may be to understand a phenomenon, or to make predictions on future observations.
- One strategy is to choose a probability distribution fits the data. Often, this involves estimating parameters.
- By viewing probability distributions as infinite hypothetical populations, techniques from survey sampling can be adapted to estimate paramaters. Same main ideas: estimate, estimator, SE, confidence interval, bias, bootstrap.

Radioactive emission

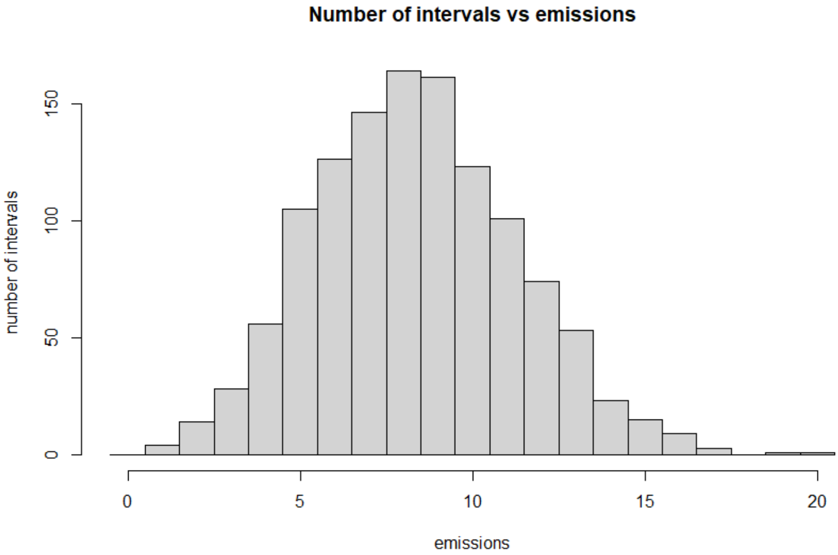
- Emissions were counted in $n = 1207$ consecutive 10-second intervals. Some frequencies are below:

Emissions	Number of Intervals
0–2	18
3	28
4	56
5	105
$\vdots$	$\vdots$
16	9
17+	5
Total	1207

The emissions have mean 8.36 and variance 8.72.

- What distribution seems to fit the histogram of emissions?

Histogram



Number of intervals vs emissions
number of intervals
0
50
100
150
0
5
10
15
20
emissions

Poisson distribution

- Let $\lambda > 0$. The Poisson(λ) random variable X has distribution

$$\Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2, \ldots$$

- Show that $\mathrm{E}(X) = \lambda$.

- Which is easier to calculate: $\mathrm{E}(X^2)$ or $\mathrm{E}\{X(X-1)\}$?

- Deduce that $\mathrm{var}(X) = \lambda$.

Fitting Poisson distribution to data

- Assume the counts $x_1, \ldots, x_n$ are realisations of IID Poisson(λ) random variables $X_1, \ldots, X_n$.
- The number of emissions per 10 seconds, λ, is a parameter. It is an unknown constant, to be estimated from data.
- What is your estimate of λ? If we think of Poisson(λ) as a population, what is its mean?
- The parameter λ is analogous to a parameter of a real population, such as μ or σ^2.

Estimating λ

- Following survey sampling, we use $\bar{x} = 8.36$ to estimate λ.
- $\bar{x}$ is a realisation of the estimator $\bar{X}$, so the SE is

$$\text{SD}(\bar{X}) = \underline{\hspace{3cm}}$$

  The estimated SE is

- An approximate 95%-CI for λ is
- We can denote the estimator by $\hat{\lambda} = \bar{X}$. So the SE is SD$(\hat{\lambda})$.

Goodness-of-fit

How well does the Poisson(8.36) distribution fit the data? The basic idea is to compare observed and expected frequencies.

The expected frequency of k emissions is

$$1207 \times \frac{8.36^k e^{-8.36}}{k!}$$

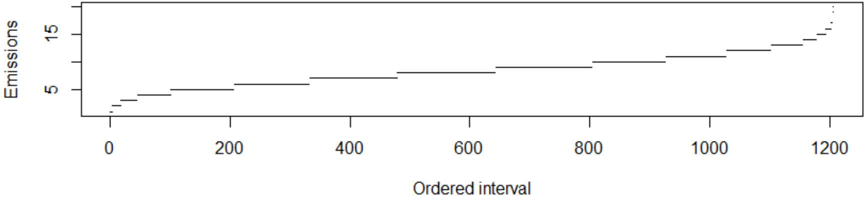
k	Observed	Expected
0–2	18	12.5
3	28	27.5
4	56	57.5
5	105	96.1
$\vdots$	$\vdots$	$\vdots$
16	9	7.7
17+	5	6.8
Total	1207	1207

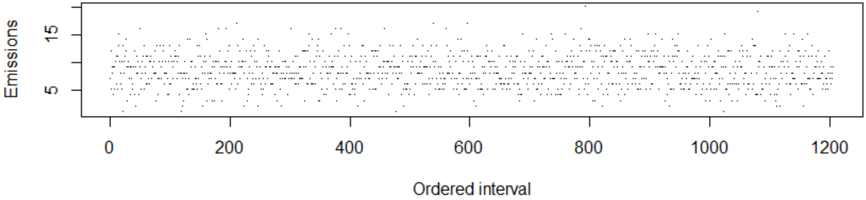
Formal methods later.

Fitting distribution to data

- Pick suitable family of distributions: Poisson, Normal, Binomial, Gamma, etc.
- **Estimation problem**: Given a distribution family, estimate parameters, estimate SE, construct CI.
- Goodness-of-fit of fitted distribution should be checked.
- It is assumed that **data are realisations of IID RV's**. This assumption should be checked, for instance by plotting data in the order of observation.

Which one looks like realisations of an RV?





Parameter space

To focus on parameters, we modify notation of mass/density function, and speak about a parameter space.
Examples:

- Poisson:
$$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, \dots,$$
$\lambda \in (0, \infty) = \mathbb{R}_+$, the parameter space.
- Bernoulli:
$$f(x|p) = p^x(1-p)^{1-x}, \qquad x = 0, 1$$
$p \in (0, 1)$, the parameter space.

Estimation problem

Assuming data $x_1, \ldots, x_n$ are realisations of IID RVs $X_1, \ldots, X_n$ with density $f(x|\theta)$, estimate θ.

The parameter θ lies in $\Theta \subset \mathbb{R}$. Θ is the parameter space.

Poisson: $\theta = \lambda$, $\Theta = \mathbb{R}_+$.
Bernoulli: $\theta = p$, $\Theta = (0, 1)$.

The statistician uses data to
(i) estimate θ
(ii) calculate an approximate SE
(iii) construct a CI (with enough data).

Two methods

- Two general ways to estimate θ from the realisations $x_1, \ldots, x_n$:
  (i) Method of moments
  (ii) Method of maximum likelihood

- In both methods, the estimate is function of data $x_1, \ldots, x_n$. Hence it is a realisation of an estimator, which is the random version of the function, of $X_1, \ldots, X_n$.

  The estimator is denoted by $\hat{\theta}$. SE is SD($\hat{\theta}$), which is estimated from data (the bootstrap).

Moments

- The k-th moment of an RV X is

$$\mu_k = \mathrm{E}(X^k), \quad k = 1, 2, \ldots$$

  Write $\mathrm{var}(X)$ in terms of the moments.

- $X \sim \mathrm{Poisson}(\lambda)$. What are the first two moments?

- $X \sim \mathrm{N}(\mu, \sigma^2)$. What are the first two moments?

Estimating moments

- Let $X_1, \ldots, X_n$ be IID with the same distribution as X. The kth sample moment is

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

  What is another way to denote $\hat{\mu}_1$? What is $\mathrm{E}(\hat{\mu}_k)$?

- We use $\hat{\mu}_k$ as an estimator of μ_k. I.e., for realisations $x_1, \ldots, x_n$,

$$\frac{1}{n} \sum_{i=1}^{n} x_i^k$$

  is an estimate of μ_k.

Method of moments: Poisson and Bermoulli

- Let $X \sim$ Poisson(λ), so $\mu_1 = \lambda$.

  MOM estimator of λ is $\hat{\mu}_1$, which is $\bar{X}$.

- Let $X \sim$ Bernoulli(p), hence $\mu_1 = p$.

  MOM estimator of p is $\hat{\mu}_1$, which is the random sample proportion $\hat{p}$.

In both examples, the MOM estimator is the same as the earlier ones, inspired by survey sampling. Are they unbiased?

MOM: normal

Let $X_1, \ldots, X_n$ be IID $\mathrm{N}(\mu, \sigma^2)$, wth parameters μ and σ^2.

- How are the parameters related to the moments of the $\mathrm{N}(\mu, \sigma^2)$ distribution?
- Obtain the MOM estimators of the parameters, in terms of $X_1, \ldots, X_n$.
- Are the MOM estimators unbiased?

Parameter space of the normal family

What are the possible values of (μ, σ^2)? This is a reasonable parameter space.

- In many applications, we are more interested in μ than σ^2. It is OK to view the parameter space as just $\mathbb{R}$.
- Then σ^2 is called a "nuisance" parameter, which has to be estimated only because we want to estimate the SE for the estimate $\bar{x}$ of μ, or to construct a CI.
- We are not so interested in the SE for the estimate of σ^2.
- But if required, will we be able to construct a 95%-CI for σ^2?

The general MOM estimator

- Let $X_1, \ldots, X_n$ be IID with density $f(x|\theta)$, where the parameter θ lies in $\Theta \subset \mathbb{R}^p$ for some $p > 1$.
- Suppose $\theta = g(\mu_1, \ldots, \mu_q)$ for some function g of the first q moments. The MOM estimator is

$$\hat{\theta} = g(\hat{\mu}_1, \ldots, \hat{\mu}_q)$$

- Write down g for Poisson, Bernoulli and normal.

Gamma distribution

- Let $X_1, \ldots, X_n$ be IID with Gamma(α, λ) distribution. $\alpha > 0$ and $\lambda > 0$ are the shape and rate parameters.

- From Tutorial 2,

$$\mu_1 = \frac{\alpha}{\lambda}, \qquad \mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$$

show that the MOM estimators are

$$\hat{\alpha} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}^2}{\hat{\sigma}^2}, \qquad \hat{\lambda} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}}{\hat{\sigma}^2}$$

Rainfall

- The amount of rain in inches was recorded in 227 storms. The mean and variance are 0.224 and 0.1338.

- Assume that the data are realisations of IID Gamma(α, λ) RV's $X_1, \ldots, X_{227}$.

  The MOM estimates of α and λ are respectively

$$\frac{0.224^2}{0.1338} \approx 0.38, \qquad \frac{0.224}{0.1338} \approx 1.67$$

SEs

- The SEs in the estimates 0.38 and 1.67

$$\mathrm{SD}(\hat{\alpha}) = \mathrm{SD}\left(\frac{\bar{X}^2}{\hat{\sigma}^2}\right), \qquad \mathrm{SD}(\hat{\lambda}) = \mathrm{SD}\left(\frac{\bar{X}}{\hat{\sigma}^2}\right)$$

  are functions of α and λ, hence unknown.

- Previously, SE is estimated by approximating the parameter by its estimate in the formula: the bootstrap.

- Here there is no formula, so an advanced bootstrap:

  Approximate the distribution by an estimated distribution.

Bootstrap approximation of SE

Let $X_1^*, \ldots, X_{227}^*$ be IID Gamma(0.38,1.67) RV's. Define

$$\bar{X}^* = \frac{1}{227}\sum_{i=1}^{227} X_i^*, \qquad \hat{\sigma}^{*2} = \frac{1}{227}\sum_{i=1}^{227}(X_i^* - \bar{X}^*)^2$$

Bootstrap approximation:

- SE of 0.38:
$$\text{SD}\left(\frac{\bar{X}^2}{\hat{\sigma}^2}\right) \approx \text{SD}\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right)$$
- SE of 1.67:
$$\text{SD}\left(\frac{\bar{X}}{\hat{\sigma}^2}\right) \approx \text{SD}\left(\frac{\bar{X}^*}{\hat{\sigma}^{*2}}\right)$$

Monte Carlo

(1) Generate realisations $x_1^*, \ldots, x_{227}^*$ from Gamma(0.38,1.67).

(2) Calculate a realisation each from $\bar{X}^{*2}/\hat{\sigma}^{*2}$ and $\bar{X}^*/\hat{\sigma}^{*2}$.

(3) Repeat (1) and (2) to get many realisations.

(4) Estimate

$$\mathrm{SD}\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right), \qquad \mathrm{SD}\left(\frac{\bar{X}^{*}}{\hat{\sigma}^{*2}}\right)$$

Back to rainfall

- Bootstrap and Monte Carlo give:

Par.	*Est.*	*SE*	*Bias*
α	0.38	0.06	0.02
λ	1.67	0.35	0.10

  Conclusion:
  α is around 0.36 ± 0.06
  λ is around 1.57 ± 0.35

- Notice the estimates have been corrected for biases. The biases are also estimated by bootstrap:

$$\mathrm{E}\left(\frac{\bar{X}^2}{\hat{\sigma}^2}\right) - \alpha \approx \mathrm{E}\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right) - 0.38, \quad \mathrm{E}\left(\frac{\bar{X}}{\hat{\sigma}^2}\right) - \lambda \approx \mathrm{E}\left(\frac{\bar{X}^{*}}{\hat{\sigma}^{*2}}\right) - 1.67$$

  then Monte Carlo.

Monte Carlo pseudocode

- Write a function that takes $(x_1, \ldots, x_n)$ and outputs $\bar{x}^2/\hat{\sigma}^2$ and $\bar{x}/\hat{\sigma}^2$.
- Use a for loop to generate the gamma realisations, feed them to the function, then store the outputs.

The for loop can be skipped in R:

- Put realisations from the gamma distribution in a matrix of many rows and 227 columns.
- Use apply().

Bootstrap

The advanced bootstrap generalises the simpler one.

- Based on $n = 1207$ Poisson(λ) realisations, λ was estimated as $\bar{x} = 8.36$.
- Advanced bootstrap: approximate Poisson(λ) by Poisson(8.36).
  Let $X_1^*, \ldots, X_n^*$ be IID Poisson(8.36) RV's.

  $$\mathrm{SD}(\bar{X}) \approx \mathrm{SD}(\bar{X}^*)$$

  But this is exactly the simpler bootstrap:

  $$\sqrt{\frac{\lambda}{n}} \approx \sqrt{\frac{8.36}{n}}$$

  Unlike the rainfall data, no Monte Carlo is needed here.

What if n gets larger?

- Monte Carlo simulation with Gamma(0.38,1.67):

	0.38		1.67	
n	*SE*	*Bias*	*SE*	*Bias*
227	0.06	0.02	0.35	0.10
2270	0.02	0.00	0.11	0.01

- As $n \to \infty$, by the Law of Large Numbers,

$$\bar{X} \to \underline{\hspace{3cm}}, \qquad \hat{\sigma}^2 \to \underline{\hspace{3cm}}$$

  Hence

$$\hat{\alpha} \to \underline{\hspace{3cm}}$$

  Similarly, bias of $\hat{\lambda}$ also goes to 0: asymptotically unbiased.

- In fact, every MOM estimator is consistent: it goes to the parameter, as $n \to \infty$.

Conclusion

- In fitting a distribution to data, parameters often need to be estimated.
- If a parameter can be expressed in terms of some moments, the MOM estimator can be defined.
- The SE and bias can be estimated by approximating the distribution by the estimated distribution (bootstrap). If no formula is available, Monte Carlo approximation is needed.
- MOM estimators can be biased, but are asymptotically unbiased. This is different from measurement and convenience sample, where the bias is not affected by sample size.
- Tutorial 4 defines moments for a real population. Which of the estimators $\bar{X}$, S^2 and $\hat{\sigma}^2$ are MOM estimators?

Statistical modeling

- The fitted distribution is called a **statistical model**.
  Poisson(8.36) is a statistical model for the 1207 radioactive emission counts.
  Gamma(0.38,1.67) is a statistical model for the 227 rain amounts.

  Bernoulli(0.5067) is a statistical model for the 10000 coin tosses of Kerrich.

- Ideally, a statistical model describes a random process: it says the data are realisations of some IID RVs.

- Fitting a distribution does not mean the process is random. Observations should be plotted in sequence to check.

- Even if the process seems random, the fitted model is an approximation. The actual values of parameters remain unknown, and are likely unknowable. Or the family of distributions might not be correct.

- George Box: "All models are wrong, but some are useful."