Point Estimation —點估計

• **What is point estimation?** *Recall. The 4 steps of statistics (Introduction, LNp.2)*


Example 6.1 (current across muscle cell membrane, TBp. 257-258)

• Bevan, Kullberg, and Rice (1979) studied random fluctuations of current across a muscle cell membrane. The cell membrane contained a large number of channels, which opened and closed at random and were assumed to operate independently. The net current resulted from ions flowing through open channels.   *cf.*   *cf.*

observed data

• They obtained 49,152 observations of the net current, $x_1, \ldots, x_{49152}$.

• The net current was the sum of a large number of roughly independent small currents.   *statistical modeling*   *CLT*   *meaning?* ← *規律*   *random variables*

• It seems appropriate to model the net current data, $X_1, \ldots, X_{49152}$ as i.i.d. $N(\mu, \sigma^2)$, where μ and σ^2 represent the mean and variance of net current. Note that the values of μ and σ^2 are unknown.   *systematic pattern*
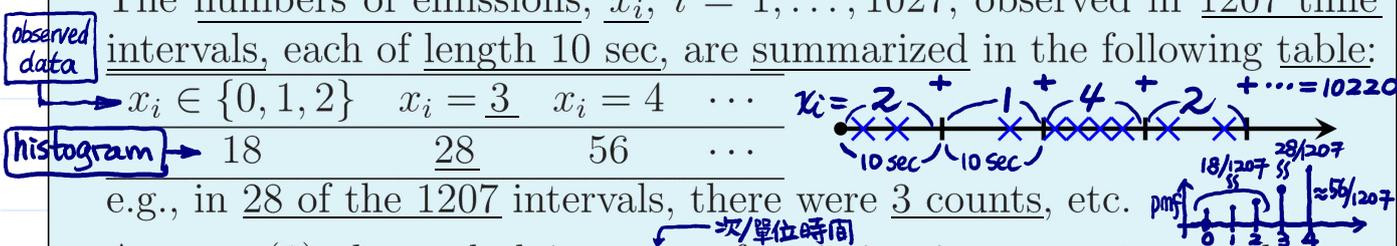
parameters

• Question: how to use the observed data, $x_1, \ldots, x_{49152}$ to gain knowledge about the values of μ and σ^2?   *transformation = ?*
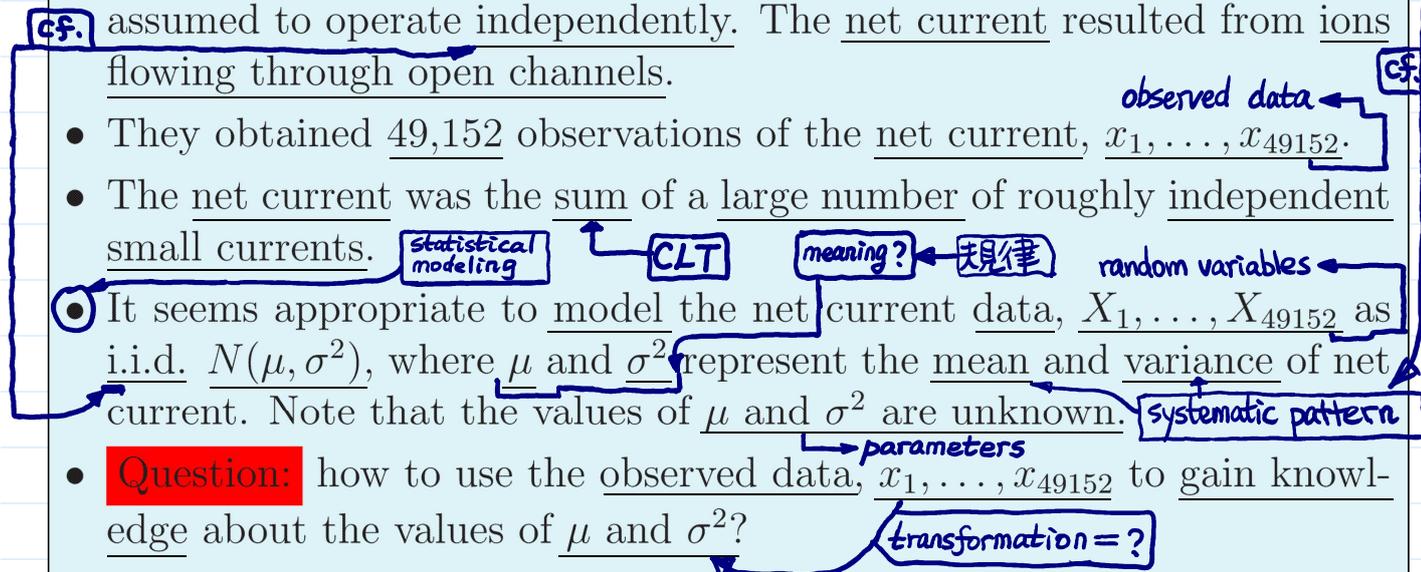
Example 6.2 (emission of alpha particles, TBp. 255-256)

• Berkson (1966) conducted an experiment about emission of alpha particles from radioactive sources. The number of emissions per unit of time is not constant but fluctuates in a random fashion.

• The experimenter recorded 10,220 times between successive emissions. The numbers of emissions, x_i, $i = 1, \ldots, 1027$, observed in 1207 time intervals, each of length 10 sec, are summarized in the following table: *observed data*

$x_i \in \{0,1,2\}$	$x_i = 3$	$x_i = 4$	$\cdots$
histogram → 18	28	56	$\cdots$

$x_i = \underbrace{2}_{10\,sec} + \underbrace{1}_{10\,sec} + 4 + 2 + \cdots = 10220$

e.g., in 28 of the 1207 intervals, there were 3 counts, etc.   *pmf* $18/1207$ $28/1207$ $\approx 56/1207$

次/單位時間

• Assume (1) the underlying rate of emission is constant over the period of observation (2) the particles come from a very large number of independent sources   *statistical modeling*   *meaning?*   *Recall. Poisson approximation to $B(n,p)$ when $n \uparrow \infty$, $p \downarrow 0$, $np \to \lambda$*   *r.v.'s*

• It seems appropriate to model the numbers of emissions $X_1, \ldots, X_{1027}$ as i.i.d. $P(\lambda)$, where λ represents the underlying rate of emission. Note that the value of λ is unknown.   *systematic pattern*   *cf.*

• Question: how to use the observed data, $x_1, \ldots, x_{1207}$, to gain knowledge about the value of λ?   *transformation = ?*
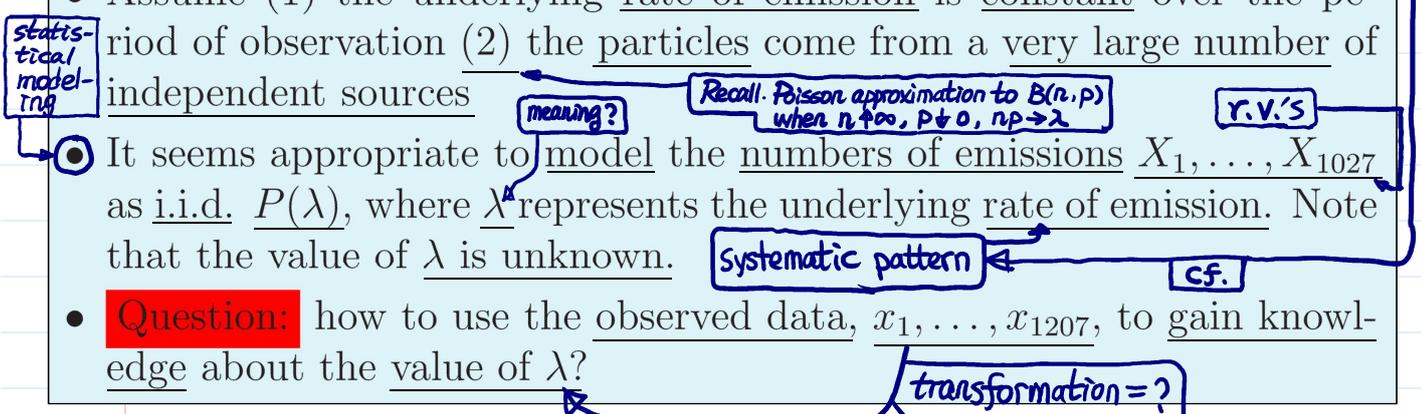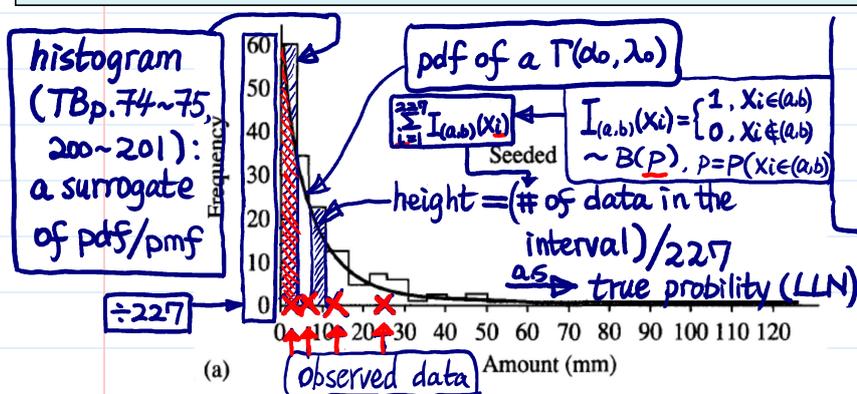
Example 6.3 (rainfall amount, TBp. 258-259)

- Le Cam and Neyman (1967) studied rainfall amounts from storms. ← *random* 直方图
- They obtained rainfall amount data, see the graphs for the histogram of the data. Let us denote $x_1, \ldots, x_{227}$ as the 227 rainfall amounts. ← *observed data*
- ⊙ The family of $\Gamma(\alpha, \lambda)$, where $\alpha > 0, \lambda > 0$, provides a flexible set of pdfs for non-negative random variable. We may model the rainfall amount data, $X_1, \ldots, X_{227}$ as i.i.d. $\sim \Gamma(\alpha, \lambda)$. Note that the values of α and λ are unknown. ← *random variables* *transformation=?* *systematic pattern*

  (*statistical modeling*)

- Question: how to use $x_1, \ldots, x_{227}$ to find a *particular* Gamma distribution $\Gamma(\alpha_0, \lambda_0)$ that can "best" fit the observed data, i.e., which pdf of Gamma is "mostly similar" to the histogram? ←*c.f.*→ Ex 6.1 & Ex.6.2

histogram (TBp.74~75, 200~201): a surrogate of pdf/pmf

pdf of a $\Gamma(\alpha_0, \lambda_0)$

$\sum_{i=1}^{227} I_{(a,b)}(X_i)$

$I_{(a,b)}(X_i) = \begin{cases} 1, & X_i \in (a,b) \\ 0, & X_i \notin (a,b) \end{cases}$

Seeded $\sim B(P), \ P = P(X_i \in (a,b))$

height = (# of data in the interval)/227

$\xrightarrow{a.s.}$ *true probability (LLN)*

÷227

Observed data

(a) Amount (mm)

Q: What information does a histogram carry?

Q: What's the difference between the 3 statistical modelings?

Ans. conceptual (Ex.6-1 & 6-2) versus empirical (Ex.6-3)

Summary (procedure of fitting a particular distribution to data, i.e. point estimation)

1. **observed data.** $x_1, \ldots, x_n$

   - **Ex 6.1:** 49152 net currents; **Ex 6.2:** 1027 numbers of emissions; **Ex 6.3:** 227 rainfall amounts

2. **statistical modeling.** Regard $x_1, \ldots, x_n$ as a realization of random variables $X_1, \ldots, X_n$, and assign $X_1, \ldots, X_n$ a joint distribution:

   呈現值

   $$\begin{array}{ll} \text{a joint cdf} & F(\cdot|\Theta), \\ \text{or a joint pdf} & f(\cdot|\Theta), \\ \text{or a joint pmf} & p(\cdot|\Theta), \end{array}$$

   This is a distribution family. ∵ Θ have many plausible values, i.e., $\Theta \in$ parameter space c.f.

   parameters

   where $\Theta = (\theta_1, \ldots, \theta_k)$, and $\theta_i's$ are **fixed constants**, but their values are **unknown**.

   Note. With the assumptions, indep.: marginal → joint; identical: accumulate information for same pattern.

   - **Ex 6.1:** i.i.d. Normal, $\Theta = (\mu, \sigma^2)$ ← $\mu \in \mathbb{R}, \ \sigma > 0$
   - **Ex 6.2:** i.i.d. Poisson, $\Theta = \lambda$ ← $\lambda > 0$
   - **Ex 6.3:** i.i.d. Gamma, $\Theta = (\alpha, \lambda)$ ← $\alpha > 0, \lambda > 0$ *a r.v.* *hat*

3. **point estimation.** Find **a function of** $X_1, \ldots, X_n$, denoted by $\widehat{\Theta}$, to estimate Θ or a function of Θ, and substitute $x_1, \ldots, x_n$ to get an estimate. *The solution to the questions in Ex.6-1~6-3.* *Why?*

Notes

1. In statistical modeling, we define

   (Objective of Data Analysis: identify systematic pattern by removing random disturb.)

   — (unknown) systematic pattern ← θ

不變	規律	certain	signal	靜	陰
變	隨機	uncertain	noise	動	陽

   — random disturbance ← r.v.'s

   in the data.

   隨機 · 萬變不離其宗 ← 不變. 規律   unknown → 規律

   • For example, unknown($\mu \in \mathbb{R}^1$)

   Say, X_i's heights

   (1) $\underline{X_1}, \ldots, \underline{X_n}$ i.i.d. $\sim \underline{N(\mu, 1)}$ known   (2) $\underline{X_1}, \ldots, \underline{X_n}$ i.i.d. $\sim \underline{N(\mu, \sigma^2)}$

   $\underline{\mu} + \varepsilon_1 \cdots \underline{\mu} + \varepsilon_n$ 規律   規律 $\mu + \sigma\varepsilon_1 \cdots \mu + \sigma\varepsilon_n$ 規律   known

   隨機 → $\varepsilon_1, \ldots, \varepsilon_n$ iid $\sim \underline{N(0,1)}$ different modeling $\varepsilon_1, \ldots, \varepsilon_n$ iid $\sim N(0,1)$

   pdf of $N(\mu, 1)$

   Information about μ & ε_i's are mixed in X_i's and indistinguable

   隨+規(K)   規(uk)

   規(uk)   All information contained in data   隨+規(K)

   $\sigma = ?$

   $\mu = ?$ $\mu = ?$
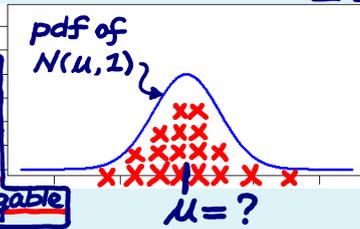
   68% → $\mu \pm \sigma$
   95% → $\mu \pm 2\sigma$
   99.7% → $\mu \pm 3\sigma$
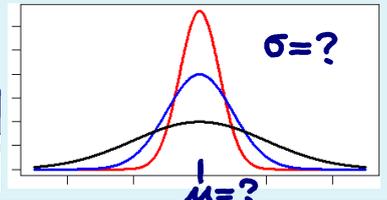
   Compare their difference:
   — prediction of μ (parameter) $\Rightarrow E(\overline{X}_n) = \mu$, $Var(\overline{X}_n) = \sigma^2/\underline{n}$
   — prediction of $\underline{X_{n+1}}$ (r.v.) $= \underline{\mu} + \underline{\epsilon_{n+1}} \Leftarrow \epsilon_{n+1} \sim N(0,1)$

   $\overline{X}_n \xrightarrow{p} \mu$ (WLLN)

   ⊙ The assumptions given in statistical modeling (i.e., joint distribution) need to be examined. ← 規(K) : $\sigma = 1$, normality, independent, $\ldots$