2. distinctions between $X_1, \ldots, X_n$ and $x_1, \ldots, x_n$ realization

- $X_1, \ldots, X_n$ are random variables while $x_1, \ldots, x_n$ are values
- different time points: before the data is collected $\Rightarrow X_1, \ldots, X_n$; after the data is collected $\Rightarrow x_1, \ldots, x_n$ enable probabilistic statements (uncertainty quantification)
- ✱ statistical inference is usually developed on the basis of $X_1, \ldots, X_n$, not $x_1, \ldots, x_n$. That is, we prefer to develop a statistical procedure that is "suitable" for all possible (future) observations under the consideration of their uncertainty, not only for a particular set of (past) observations. $\hat{\theta} \Leftrightarrow F(\cdot|\hat{\theta})$ 決定身處的平行世界 i.e. point estimation

e.g. Find $\hat{\theta}$

conceptual model

3. (TBp.259) reasons for fitting a particular distribution to data

Ex.6-1, Ex.6-2

- Scientific theory may suggest the form of a probability distribution, and parameters of that distribution may be of direct interest.

usually meaningful cf.

- For descriptive purposes as a method of data summary or compression. → e.g. in Ex.6-3, data (of dimension 227) is reduced to 2 parameters

Ex.6-3 empirical model

- A probability model may play a role in complex modeling. For example, utility companies may model daily temperatures as random variables from a distribution, which may be used in simulations of effects of various pricing and generation schemes.

systematic patterns

∈ parameter space = the collection of all possible values of Θ.

Definition 6.1 (parameter)

The fixed but unknown constant(s), i.e., Θ, in the joint distribution of $X_1, \ldots, X_n$ are called **parameters.** Sometimes, functions of Θ are also called parameters.

both are often called 'estimate' in Textbook

Definition 6.2 (statistic, estimator, estimate, sampling distribution, TBp.260)

統計量

- A **statistic** is a function of $X_1, \ldots, X_n$. (no parameters in the function) $\bar{X}_n \overset{cf.}{\longleftrightarrow} \bar{X}_n - \mu$ (a r.v.)

估計式

- A (point) **estimator** $\hat{\theta}$ of a parameter θ is a statistic used to estimate θ, and a (point) **estimate** is a value of $\hat{\theta}$ computed based on the observed data $x_1, \ldots, x_n$. Note that an estimator is a random variable and an estimate is a number.

估計值

depends on Θ (unknown)

- The distribution of an estimator is called **sampling distribution.**

Definition 6.3 (standard error, estimated standard error, TBp. 262)

- The **standard error** of an estimator is the standard deviation of its sampling distribution, i.e., $\sqrt{Var_\theta(\hat{\theta})}$. ← a function of parameters ← estimate
- An estimate of the standard error is called **estimated standard error.**

$E_\theta(\hat{\theta}-\theta)$ ‖ average error

Definition 6.4 (bias, unbiased estimator, TBp. 262)

- The **bias** of an estimator is defined as $E_\theta(\hat{\theta}) - \theta$. ← a function of parameters
- An estimator is called **unbiased** if $E_\theta(\hat{\theta}) = \theta$, i.e., bias equals zero.

不偏

❖ **Reading**: textbook, 8.1, 8.2, 8.3

• Method of finding estimators I --- method of moments

Recall. The kth moment of a distribution, if exists, is

$\mu_k(\theta)$: a function of parameters → $\mu_k \equiv E_\theta(X^k)$, parameters (unknown)

where X is a random variable following that distribution. function of data

Definition 6.5 (sample moment, TBp. 260)

If $X_1, X_2, \ldots, X_n$ are i.i.d. random variables from a distribution, the kth **sample moment** is defined as:

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

Let $Y_i = X_i^k \Rightarrow Y_1, \cdots, Y_n$: i.i.d.
$\hat{\mu}_k = (Y_1 + \cdots + Y_n)/n \xrightarrow{LLN} E(Y_i) = E(X_i^k) = \mu_k$

Note. $\hat{\mu}_k$ is an estimator of μ_k and $\hat{\mu}_k$ is unbiased, i.e. $E(\hat{\mu}_k) = \mu_k$.

Definition 6.6 (method of moments, TBp. 261)

Step 1. Calculate low-order moments, find expressions for the moments in terms of the parameters.

Step 2. Invert the expressions found in Step 1, finding new expressions for the parameters in terms of the moments.

Step 3. Insert the sample moments into the expressions obtained in Step 2, thus obtaining estimator of the parameters.

For example, suppose θ_1, θ_2 are parameters such that

usually, # of lower-order moments needed = # of parameters

$\mu_1 = h_1(\theta_1, \theta_2)$, $\mu_2 = h_2(\theta_1, \theta_2)$ $\Rightarrow$

$$\theta_1 = g_1(\mu_1, \mu_2); \qquad \theta_2 = g_2(\mu_1, \mu_2)$$

then the method of moments estimators of θ_1 and θ_2 are

a function of data $X_1, \cdots, X_n$

$$\hat{\theta}_1 = g_1(\hat{\mu}_1, \hat{\mu}_2); \qquad \hat{\theta}_2 = g_2(\hat{\mu}_1, \hat{\mu}_2).$$

Example 6.4 (Poisson distribution, TBp. 261)

- Suppose that $X_1, \cdots, X_n$ are i.i.d. $\sim P(\lambda)$. Then the first moment of $P(\lambda)$ is λ. The first sample moment is

  $\boxed{\mu_1 = \lambda}$

  $$\hat{\mu}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

  Therefore, the method of moment estimator of λ is $\hat{\lambda} = \overline{X}$.

  reasonable ?

- As an example, asbestos fibers on filters were counted as part of a project to develop measurement standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and punches of 3-mm diameter were taken from the filter. Counts of the numbers of fibers in each of 23 grid squares are

  $X_1, \cdots, X_{23} \overset{iid}{\sim} P(\lambda)$

  statistical modeling

  $x_1, \cdots, x_{23}$

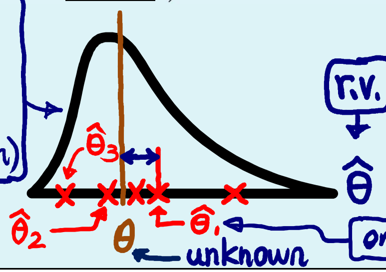
31	29	19	18	31	28	34	27	34	30	16	18
26	27	27	18	24	22	28	24	21	17	24	

  and the method of moments estimate of λ is $\hat{\lambda} = 24.9$.

  Q. Is this answer enough?

- **Questions:**
  1. Is the estimate (i.e., 24.9) the real λ? ← Note. $\hat{\lambda}$ (estimator) is a r.v.
  2. Next time when the same procedure (same sample size, same estimator, same ...) is repeated again to get a new estimate, ★ how far the future estimate will be away from 24.9? → need an uncertainty quantification.
  3. In which range will you expect say 95% of the future estimate falls? (e.g., [24.8, 25] or [15, 35]?)
  4. This is a question related to the stability/uncertainty/variation of the estimator. (or $\hat{\lambda}-\lambda$) ← $\boxed{\text{a r.v.}}$
  5. How to characterize the stability of an estimator? (**Note.** We have to answer the question using only the observed data.) ← i.e., not knowing the true value of λ.

$\boxed{\text{pdf of the sampling dist. of } \hat{\theta} \text{ (unknown)}}$



To evaluate the stability/uncertainty of an estimation procedure, it is required to know what the sampling distribution is.

$\boxed{\text{only observe this}}$ └ to estimate

Example 6.5 (cont. Ex.6.4, TBp. 262)

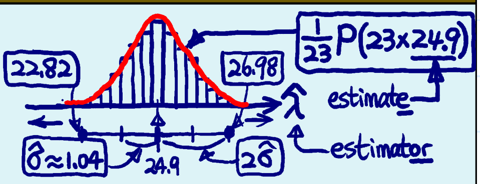
- **Exact sampling distribution** of $\hat{\lambda}$: ← $\boxed{\bar{X}_n}$

Because $X_1, X_2, \ldots, X_n$ i.i.d. $\sim P(\lambda)$,

$$n\hat{\lambda} = S = \sum_{i=1}^{n} X_i \boxed{\sim P(n\lambda)} \quad \text{unknown}$$

$$\boxed{\hat{\lambda} = \frac{S}{n}} \qquad E(\hat{\lambda}) = \frac{1}{n}E(S) = \lambda, \quad \mathrm{Var}(\hat{\lambda}) = \frac{1}{n^2}\mathrm{Var}(S) = \lambda/n\,.$$

$\boxed{\text{true } \lambda \text{ is unknown}}$



$\boxed{\text{meaning: } Z \sim P(n\lambda), \text{ dist. of } \frac{1}{n}Z}$

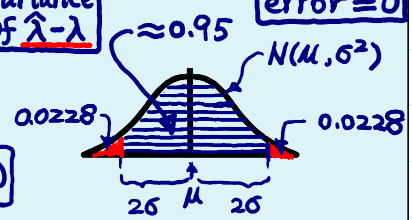
Thus, (1) the sampling distribution of $\hat{\lambda}$ is $\frac{1}{n}P(n\lambda)$, (2) $\hat{\lambda}$ is unbiased, $\boxed{\text{average error} = 0}$ and (3) the standard error of $\hat{\lambda}$ is $\sigma_{\hat{\lambda}} = \sqrt{\lambda/n}$. ← $\boxed{\text{variance of } \hat{\lambda}-\lambda}$

$\boxed{\text{a function of unknown parameter } (\lambda)}$

- Estimated standard error of $\hat{\lambda}$ is

$\boxed{\text{assume } \hat{\lambda}=24.9 \text{ is the true value of } \lambda}$

$$s_{\hat{\lambda}} = \sqrt{\hat{\lambda}/n} = \sqrt{24.9/23} = 1.04.$$

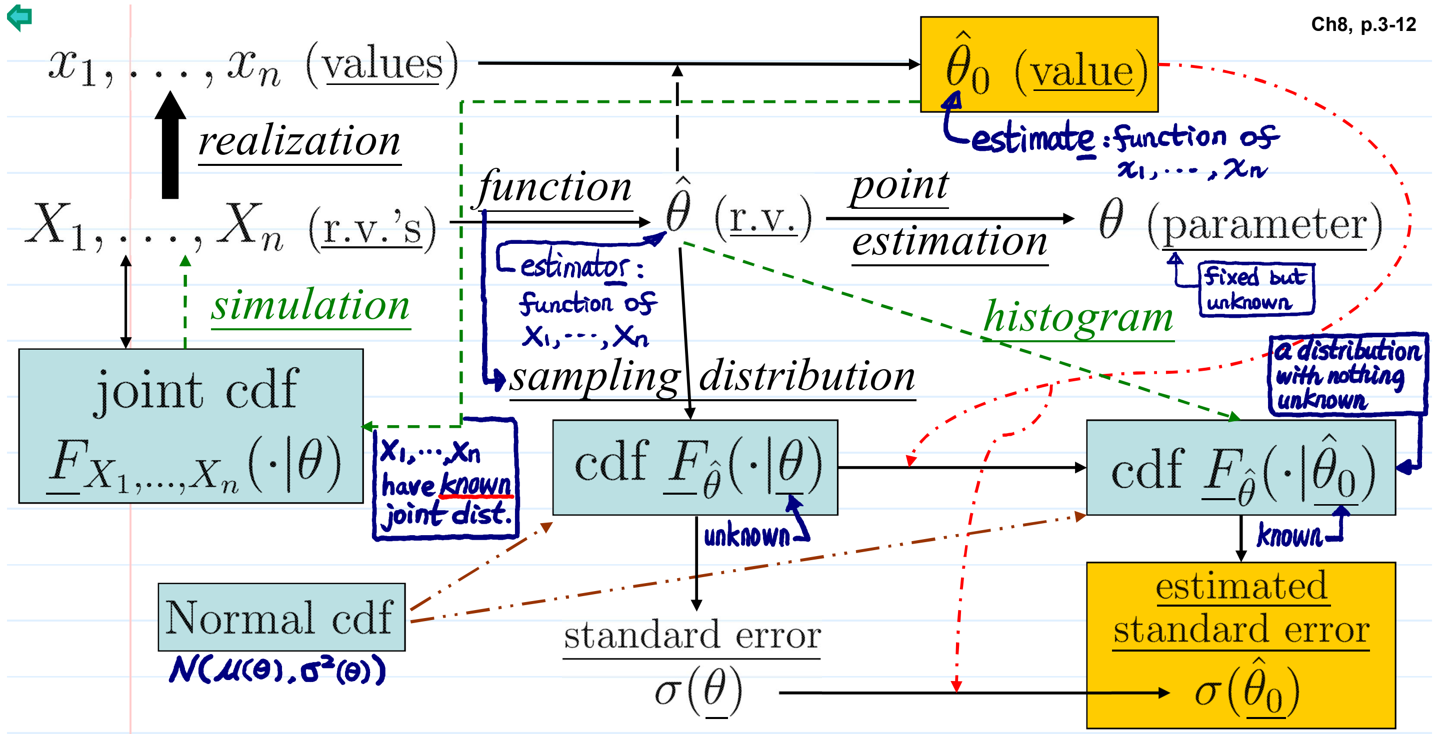


- **Asymptotical method**: By CLT, sampling distribution of $\hat{\lambda}$ is approximately normal when n is large enough. Since a normally distributed random variable is very unlikely to be more than 2 standard deviation away from its mean, errors in $\hat{\lambda}$ is very unlikely to be more than 2.08.

$\boxed{S \xrightarrow{d} N(n\lambda, n\lambda) \quad \hat{\lambda} \xrightarrow{d} N(\lambda, \tfrac{\lambda}{n})}$

$\boxed{P(|\hat{\lambda}-\lambda| > 2.08) \approx 0.0458}$

$24.9 \pm 2 \times 1.04 = [22.82, 26.98]$



- exact distribution $\Rightarrow$ the form of $F_{\hat\theta}(\cdot|\theta)$ is known
- asymptotical method $\Rightarrow$ the form of $F_{\hat\theta}(\cdot|\theta)$ is close to Normal cdf (usually when n is large)
- simulation method $\Rightarrow$ useful when the form of $F_{\hat\theta}(\cdot|\theta)$ is unknown

Exercise: Identify the elements in the diagram for each examples.

Example 6.6 (Normal distribution, TBp. 263)

- The first and second moments for $N(\mu, \sigma^2)$ are

$$\begin{cases} \mu_1 = E(X) = \mu \\ \mu_2 = E(X^2) = \mu^2 + \sigma^2 \end{cases}$$

$\because Var(X) = E(X^2) - [E(X)]^2$

$$\Rightarrow \begin{cases} \mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2 \end{cases}$$

statistical modeling

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim N(\mu, \sigma^2)$, then the method of moment estimators of μ and σ^2 are

reasonable?

$$\begin{cases} \hat\mu = \overline{X} \\ \hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2 = \frac{1}{n}\overbrace{\sum_{i=1}^n (X_i - \overline{X})^2}^{=\sum_{i=1}^n X_i^2 + n\overline{X}^2 - 2n\overline{X}^2} \end{cases}$$

$\sim \chi^2_{n-1}$

cf. (LN, CH1~6, p.80 & 94) $\hat\sigma^2 \xrightarrow{P} \sigma^2$, $\hat\sigma^2 = \frac{n-1}{n} S_n^2 = \frac{\sigma^2}{n} \boxed{\frac{(n-1)S_n^2}{\sigma^2}}$

By Thm. 4.1 LNp.1~6, p.80

- Sampling distribution of $\overline{X}$ is Normal$(\mu, \frac{\sigma^2}{n})$ and sampling distribution of $\hat\sigma^2$ is $\frac{\sigma^2}{n}\chi^2_{n-1}$. Furthermore, $\overline{X}$ and $\hat\sigma^2$ are independent.

$Var(\hat\sigma^2) = \frac{\sigma^4}{n^2} 2(n-1) = \frac{2\sigma^4}{n} \frac{n-1}{n}$

Example 6.7 (Gamma distribution, TBp. 263-264)

- The first two moments of the $\Gamma(\alpha, \lambda)$ are

$\sigma^2 = Var(X_1)$

$\mu_2 = \sigma^2 + \mu^2$

$$\begin{cases} \mu_1 = \alpha/\lambda \\ \mu_2 = \alpha(\alpha+1)/\lambda^2 \end{cases} \Rightarrow \begin{cases} \lambda = \mu_1/(\mu_2 - \mu_1^2) \\ \alpha = \lambda\mu_1 = \mu_1^2/(\mu_2 - \mu_1^2) \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be i.i.d. $\sim \Gamma(\alpha, \lambda)$, then the method of moment estimators of λ and α are

statistical modeling

$$\hat{\lambda} = \overline{X}/\hat{\sigma}^2, \qquad \hat{\alpha} = \overline{X}^2/\hat{\sigma}^2$$

where $\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2$. $= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ (LNp.13)

- As a concrete example, let us consider the fit of the rainfall amounts during 227 storms in Illinois from 1960 to 1964 (the data, listed in Problem 42, Textbook p.414). For the data, $\overline{X} = .224, \hat{\sigma}^2 = .1338$, therefore $\hat{\alpha} = .375$ and $\hat{\lambda} = 1.674$. estimate cf.

- What are the sampling distributions of $\hat{\alpha}$ and $\hat{\lambda}$?
  - Exact distributions are complicated.
  - Asymptotic method: Use central limit theorem and other asymptotic theorems to obtain the limiting distributions.
  - simulation method: bootstrap. ← a Monte Carlo method

Definition 6.7 (parametric bootstrap, TBp. 264-265) cf. non-parametric (# of parameters = ∞)

pseudo random numbers

i.e. do transformation

② Generate many, many samples of size n from a distribution. (True values of parameters in the distribution are unknown, so ① use their estimates). From each of the samples, ③ compute the estimates of parameters. The ④ histogram of these estimates should give a good idea of the ⑤ sampling distribution of estimator.

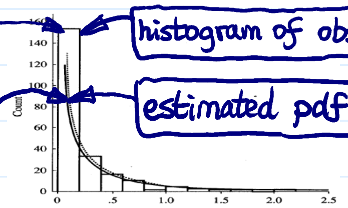
check the graph in LNp.12

non-parametric bootstrap

histogram of observed data

parametric bootstrap

estimated pdf of $X_1, \cdots, X_n$



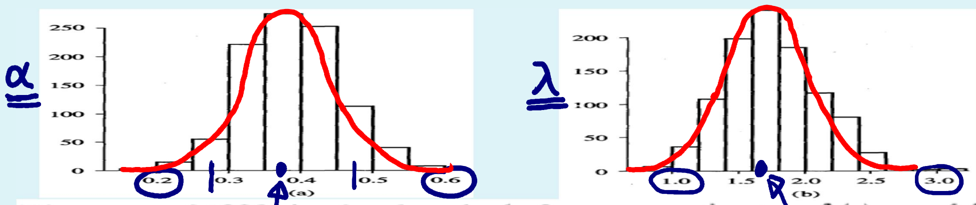
replace the true distribution of $X_1, \cdots, X_n$ (unknown) by the estimated distribution (known)

regard empirical cdf as the true cdf.

Example 6.8 (cont. Ex.6.7, bootstrap, TBp. 265-266)

Generate 1000 samples of size 227 from $\Gamma(\hat{\alpha}, \hat{\lambda})$, where $\hat{\alpha}, \hat{\lambda}$ are set to be .375 and 1.674, respectively. From each of 1000 samples, compute the estimates of α and λ using the estimators $\hat{\lambda} = \overline{X}/\hat{\sigma}^2$, $\hat{\alpha} = \overline{X}^2/\hat{\sigma}^2$. Denote the 1000 estimates by α_i^*, λ_i^*, $i = 1, \ldots, 1000$. Histograms of them indicate the variability that is inherent in estimating the parameters from a sample of this size.

surrogate of $F_\theta(\cdot|\hat{\theta}_0)$.



Histogram of 1000 simulated method of moment estimates of (a) α and (b) λ.

0.375 1.674

We can find from the 1000 α_i^*'s and λ_i^*'s that: large enough.

- The histograms looks like normal. (sample size = 227, not 1000)
- The histogram suggests that if $\alpha = 0.375$, then it is not very unusual that $\hat{\alpha}$ is in error by 0.1 or more.
- The histograms are centered at 0.375 and 1.674, the (regarded-as-true) parameter values used in the simulation.
- Estimated standard error of $\hat{\alpha}$ is

The estimation procedure is effective

may be unbiased or almostly unbiased

assume the estimate is (or close to) the true value of parameter

$$s_{\hat{\alpha}} = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\alpha_i^* - \overline{\alpha})^2} = 0.06,$$

equivalent to the standard deviation of the histogram

$\pm 2 s_{\hat{\alpha}} = \pm 0.12$

where $\overline{\alpha}$ is the mean of the 1000 values. Also, in the same way, $s_{\hat{\lambda}} = 0.34$.

❖ **Reading**: textbook, 8.4