Applications in Parameter Estimation / Statistical Inference

So far, we have studied properties of r.v.'s, probability models etc. However, we have assumed prior knowledge of the probability model which governs the outcomes of an experiment. In practice, however, we encounter numerous situations where the probability model is not known in advance & we collect data from experiments to learn the model. This area is known as "Statistical Inference", which governs the use of measurements to discover the properties of a probability model.

Let us return to an experiment (governed by some unknown probability model); $\rightarrow f_X(x)$.
We conduct n independant experiments.
$\rightarrow X_1 \; X_2 \ldots$ (each of them are independent & identically distributed).

$$M_n = \frac{X_1 + X_2 \cdots + X_n}{n}$$ is an estimate of the mean of X.

(Sample mean)

The sample mean is an example of a **Point Estimate**, which is a single number that is as close as possible to the parameter to be estimated.

Another type of estimate is a <u>confidence interval estimate</u> which is a range of numbers that contain the parameter with high probability.

3 Key Properties of Point Estimates

Bias | **Consistency** | **Accuracy**

Suppose we want to estimate a parameter r of $f_X(x)$ $\left\{ \text{eg} \quad \begin{array}{l} r \to \text{mean or} \\ r \to \text{variance etc..} \end{array} \right\}$.

We conduct independent experiments

$$\underbrace{\underbrace{X_1}_{\hat{R}_1} \quad X_2}_{\hat{R}_2} \quad X_3 \quad \cdots\cdots\cdots$$

$\hat{R}_2 \quad \cdots \quad \hat{R}_n \to$ some estimator of r (function of $X_1, \ldots, X_n$).

① $\hat{R}$ is unbiased estimator of r if

$$E[\hat{R}] = r.$$

For Unbiased estimator:Bias=(I

② A sequence of estimates $\hat{R}_1\ \hat{R}_2 \ldots$ of parameter r is consistent if for any $\epsilon > 0$

$$\lim_{n\to\infty} P(|\hat{R}_n - r| \geq \epsilon) = 0$$

or $\hat{R}_n \xrightarrow{\text{in Probability}} r.$

[Asymptotically unbiased:
A sequence of estimators $\hat{R}_n$ of r is asymptotically unbiased if

$$\lim_{n\to\infty} E[\hat{R}_n] = r.$$

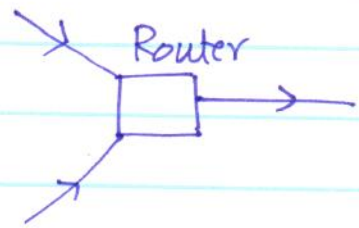
③ MSE of estimator $\hat{R}$ of parameter r is

error $$e = E[(\hat{R}-r)^2].$$

Note: if $\hat{R}$ is an unbiased estimator, i.e $E(\hat{R}) = r$, then MSE, e is simply the Variance of $\hat{R}$.

(4) If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ has MSE $e_n = Var(\hat{R}_n)$ with $\lim_{n \to \infty} e_n = 0$, then sequence $\hat{R}_n$ is consistent.

Example 1.



In an interval of k seconds, the number N_k of packets passing through an internet router is a Poisson r.v. with expected value $E[N_k] = k\gamma$ packets. Let $\hat{R}_k = \frac{N_k}{k}$ denote an estimate of γ.

(1) Is $\hat{R}_k$ an unbiased estimate of γ?

Yes $E(\hat{R}_k) = E\left(\frac{N_k}{k}\right) = \frac{1}{k} E(N_k) = \frac{k\gamma}{k} = \gamma$ ✓

(2) What is MSE e_k of estimate $\hat{R}_k$?

$e_k = E((\hat{R}_k - \gamma)^2) = \frac{k\gamma}{k^2} = \frac{\gamma}{k}$. $\Rightarrow e_k \to 0$ as $k \to \infty$.

(3) Is the sequence of estimators $\{\hat{R}_k\}$ consistent?

Yes $\hat{R}_k \xrightarrow{m.s.} \gamma \Rightarrow \hat{R}_k \xrightarrow{\text{in Probability}} \gamma$.

$\Rightarrow \hat{R}_k$ is consistent.

$\boxed{\text{Lec. 29, 30 Nov 4, Nov. 6}}$

If we use M_n, the sample mean to estimate μ, then using Chebyshev's inequality,

$$P(|M_n - \mu| \geq c) \leq \frac{Var(X)}{nc^2} = \alpha$$

or

$$P(|M_n - \mu| < c) \geq 1 - \frac{Var(X)}{nc^2} = 1 - \alpha$$

$|M_n - \mu| \geq c \Rightarrow$

$$\boxed{M_n - c \leq \mu \leq M_n + c}$$

$$\boxed{2c \Rightarrow \text{Confidence Interval}}$$

$$\boxed{1 - \alpha \Rightarrow \text{Confidence Coefficient.}}$$

If α is small, we are highly confident that M_n is in the interval $(\mu - c, \mu + c)$.

In a practical application

* c indicates the desired accuracy
* α indicates our confidence that we have achieved this accuracy.
* n tells us how many samples do we need to achieve this accuracy.

Eg Suppose we perform n independent trials of an experiment and use the relative frequency $\hat{P}_n(A)$ to estimate $P(A)$. Use the chebyshev ineq. to calculate the smallest n such that $\hat{P}_n(A)$ is in a confidence interval of length 0.02 with confidence 0.999.

$\hat{P}_n(A) =$ Sample mean of the indicator r.v. X_A

$$X_A = \begin{cases} 1 & \text{if event } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

$E(X_A) = P(A)$

$Var(X_A) = P(A)(1-P(A))$ $\quad \Rightarrow X_A \sim Ber(P(A))$

$E(\hat{P}_n(A)) = P(A)$.

$$\Rightarrow \quad P\{ |\hat{P}_n(A) - P(A)| < c] \geq 1 - \frac{P(A)(1-P(A))}{n c^2}$$

Note that

$P(1-P) \leq 0.25$ for any $P \in [0,1]$

$\Rightarrow \quad P(A)(1-P(A)) \leq \frac{1}{4}$ for any $P(A)$

$$\Rightarrow \quad P(|\hat{P}_n(A) - P(A)| < c) \geq 1 - \frac{1}{4nc^2}.$$

Confidence interval $= 2c = 0.02$

$\Rightarrow c = 0.01$

confidence $= 1-\alpha = 0.999$

$$\Rightarrow \quad 1 - \frac{1}{4nc^2} \geq 0.999$$

$$\Rightarrow \quad 1 - \frac{1}{4n(0.01)^2} \geq 0.999$$

$$\Rightarrow \quad n \geq 2.5 \times 10^6 \text{ trials!}$$

Using the CLT for Confidence Estimates

Recall, the CLT says that

$$\frac{M_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1)$$

as $n \to \infty$.

$\Rightarrow$ For large enough n,

$$P(M_n - c \le \mu \le M_n + c)$$

$$= P(-c \le (M_n - \mu) \le c)$$

$$= P\left(-\frac{c\sqrt{n}}{\sigma} \le \left(\frac{M_n - \mu}{\sigma/\sqrt{n}}\right) \le \frac{c\sqrt{n}}{\sigma}\right)$$

$$= 1 - 2P\left(\frac{M_n - \mu}{\sigma/\sqrt{n}} \ge \frac{c\sqrt{n}}{\sigma}\right)$$

$$= 1 - 2Q\left(\frac{c\sqrt{n}}{\sigma}\right)$$

$$= 1 - \alpha$$

$$\Rightarrow \quad \alpha = 2Q\left(\frac{c\sqrt{n}}{\sigma}\right) = 2\left(1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right)$$

Example: $X_1, X_2 \ldots$ is a sequence of iid exponential r.v.'s with expected value 5.

(a) What is $Var(M_9)$, the variance of the Sample mean based on nine trials?

$X_1, X_2 \ldots$ exponential $\Rightarrow F_X(x) = \begin{cases} 1 - e^{-x/5}, & x \geq 0 \end{cases}$

$\sigma^2 = 25$

$$Var(M_9) = \frac{\sigma^2}{9} = 25/9$$

(b) What is $P(X_1 > 7)$?

$$P(X_1 > 7) = 1 - P(X_1 \leq 7) =$$
$$= 1 - (1 - e^{-7/5}) = e^{-7/5} \approx 0.247$$

(c) What is. $P(M_9 > 7)$? Find estimate for it using CLT.

$$\cancel{P(M_9 > 7) = 1 - P(M_9 \leq 7)}$$
$$\cancel{= 1 - P\left(\frac{M_9 - \mu}{\sigma/\sqrt{n}} \leq \frac{7-\mu}{\sigma/\sqrt{n}}\right)}$$
$$\cancel{= 1 - \Phi(}$$

$$P(M_9 > 7) = 1 - P(M_9 \le 7)$$

$$= 1 - P\left(\frac{M_9 - \mu}{\sigma/\sqrt{9}} \le \frac{7-\mu}{\sigma/\sqrt{9}} \right).$$

$$= 1 - \Phi\left(\frac{7-\mu}{\sigma/\sqrt{9}} \right)$$

$$= 1 - \Phi\left(\frac{7-5}{5 \times \frac{1}{3}} \right)$$

$$\simeq 1 - \Phi\left(\frac{6}{5} \right) = 0.1151$$

Example. A telephone call is either voice / data

$$P(\text{Voice Call}) = 0.8$$
$$P(\text{Data Call}) = 0.2$$

~~Data & voice~~ calls occur independently of each other.

K_n = # of data calls in a collection of n phone calls

(a) $E(K_{100}) \overset{?}{=}$ Expected # of data calls in 100 calls?

$$D = \begin{cases} 1 & \text{w.p. } 0.2 \\ 0 & \text{w.p. } 0.8 \end{cases}$$

$D = 1$ signifies a data call.

$$E(D) = 0.2$$
$$Var(D) = 0.2 - (0.2)^2 = 0.16$$

$$E(K_{100}) = E(D_1 + D_2 + \ldots + D_{100}) = 100 \times 0.2 = 20$$

(b) $Var(K_{100}) = \sqrt{100\, Var(D)} = \sqrt{16} = 4$

(c) $P(K_{100} \geq 18) = P\left(\frac{K_{100} - 20}{4} \geq \frac{18-20}{4}\right)$

$$\underset{(CLT)}{\cong} 1 - \Phi\left(\frac{18-20}{4}\right) = 1 - \Phi\left(-\tfrac{1}{2}\right)$$
$$= \Phi(1/2)$$
$$= 0.6915$$

(d) $P(16 \le K_{100} \le 24)$

$$\underset{\text{CLT}}{\cong} \Phi\left(\frac{24-20}{4}\right) - \Phi\left(\frac{16-20}{4}\right)$$

$$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1$$

$$= 0.6826$$