Systems and Signals 344

2025

Lecture 25

Laws of Large Number &
Point Estimates of Model Parameters

Section 10.3-10.4

Recap: Markov inequality

The **Markov inequality** states that for a random variable X, such that $P(X < 0) = 0$ (taking on nonnegative values only), and a constant $c > 0$:

$$P(X \geq c) \leq \frac{\mathrm{E}[X]}{c}$$

This gives the *upper bound* of the probability. And note that the distribution of X is not important.

Recap: Chebyshev Inequality

The **Chebyshev Inequality** states that for an arbitrary random variable X and constant $c > 0$

$$P(|X - \mu_X| \geq c) \leq \frac{\text{Var}[X]}{c^2}$$

The Chebyshev inequality states that the difference between X and $\mathbf{E}[X]$ is somehow limited by $\mathbf{Var}[X]$.

It provides the upper bound of the *probability* that the value of X lies outside some margin c about the expected value.

The Chebyshev inequality generally provides a tighter bound than the Markov inequality. In particular, when the variance of X is very small.

Recap: Sample Mean

We can consider sampling as n trials, producing n iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean of X is the RV:

$$M_n(X) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mathbf{E}[M_n(X)] = \mathbf{E}[X]$$

$$\mathbf{Var}[M_n(X)] = \frac{\mathbf{Var}[\mathrm{X}]}{n}$$

Note that this last one is NOT the sample variance, it is the variance of the sample mean.

Law of Large Numbers

The sample mean $M_n(X)$ converges to $\mathbf{E}[X]$ and the relative frequency of event A converges to $P(A)$ as n, the number of independent trials of an experiment, increases without bound.

Law of Large Numbers

The **Weak Law of Large Numbers (Finite Samples)** states that for any constant $c > 0$

$$\textbf{1. } P(|M_n(X) - \mu_X| \geq c) \leq \frac{\text{Var}[X]}{nc^2}$$

$$\textbf{2. } P(|M_n(X) - \mu_X| < c) \geq 1 - \frac{\text{Var}[X]}{nc^2}$$

In words, this states that the probability that the *sample mean* is more than c units from $\mathbf{E}[X]$ can be made arbitrarily small by letting the number of samples n become large.

If $n \to \infty$, we obtain the infinite limit result in the next theorem.

Law of Large Numbers

The **Weak Law of Large Numbers (Infinite Samples)** states that if X has finite variance, for any constant $c > 0$

1. $\lim_{n \to \infty} P(|M_n(X) - \mu_X| \geq c) = 0$
2. $\lim_{n \to \infty} P(|M_n(X) - \mu_X| < c) = 1$

And, per definition:

$$\lim_{n \to \infty} P(|M_n(X) - \mu_X| \geq c) = 1 - \lim_{n \to \infty} P(|M_n(X) - \mu_X| < c)$$

This type of convergence is an example of **convergence in probability**, since the probability that the sample mean differs from the true mean by more than an arbitrarily small value vanishes.

Thus, the sample mean approaches the true mean with probability 1.

Law of Large Numbers

If there is a *weak* law of large numbers, then there is also a *strong* law of large numbers.

The **strong law of large numbers** states that *with probability* $\mathbf{1}$, the sequence $M_1, M_2, \ldots$ has the limit μ_X.

We don't cover this, but the difference between the strong law and the weak law of large numbers is subtle and rarely arises in practical applications of probability theory.

Law of Large Numbers

The weak law of large numbers also validates the **relative frequency** interpretation of probability.

Choosing X so that $P(A) = \mathrm{E}[X]$

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

If $\hat{P}_n(A) = M_n(X_A) = \dfrac{X_{A1} + X_{A2} + \ldots + X_{An}}{n}$ then,

$\hat{P}_n(A)$ converges to $P(A)$ as $n \to \infty$

$$\lim_{n\to\infty} P(|\hat{P}_n(A) - P(A)| \geq c) = 0$$

Point Estimates of Model Parameters

$\hat{R}$, an estimate of a parameter, r, of a probability model is unbiased if $\mathbf{E}[\hat{R}] = r$.

A sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *is consistent if* $\lim_{n \to \infty} \hat{R} = r$.

The sample mean is an unbiased estimator of μ_X.
The sequence of sample means is consistent.
The sample variance is a biased estimator of $\mathbf{Var}[X]$.

Point Estimates of Model Parameters

- Now we shift our focus to the experiments performed in order to obtain information about a probability model.
- We will only look at estimating models based on the *expected value* and the *variance* of a random variable, but the subject is much broader than that.
- The general problem is the estimation of a *parameter* ($\boldsymbol{r}$) of a probability model which is any number that can be calculated from the probability model.
    - For example, the probability of arbitrary event A ($P(A)$) is a model parameter.
    - Or, for example, the mean of a univariate distribution or the correlation coefficient of a bivariate distribution
- We will focus on a **point estimate**, a number that is as close as possible to the parameter to be estimated.

Point Estimates of Model Parameters

Before presenting estimation methods based on the sample mean, we introduce three properties of point estimates: **bias**, **consistency**, and **accuracy**.

Once we have defined these terms, we then develop the theorems.

Consistency

Consider the indefinite *iid* sample values $X_1, X_2, \ldots$ and assume r is a parameter of the probability model with estimates $\hat{R}_1, \hat{R}_2, \ldots$ which are also random variables, such that

$$\hat{R}_1 = g(X_1), \hat{R}_2 = g(X_1, X_2), \ldots, \hat{R}_n = g(X_1, X_2, \ldots, X_n).$$

The function simply indicates that we use the sample values to calculate the parameter of interest.

When the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *converges* in probability to parameter r, we say the estimator is **consistent**. This happens if for any $\epsilon > 0$,

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

Bias

In statistics, the **bias** of an estimator is the difference between this estimator's *expected value* and the *true value* of the parameter being estimated.

An estimate $\hat{R}$, of parameter r is **unbiased** if $\mathbf{E}[\hat{R}] = r$;
otherwise $\hat{R}$ is **biased**.

It could be that sometimes $\hat{R} > r$ and other times $\hat{R} < r$, but it is desirable to have the typical value of $\hat{R}$ to be equal to r. Therefore, unbiased estimates are preferred.

Bias is concerned with only a single estimator. If you are interested in a sequence of estimators...

Asymptotically Unbiased Estimator

The sequence of estimators $\hat{R}_n$ of parameter r is **asymptotically unbiased** if

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

Now it is only unbiased for $n \to \infty$

Mean Square Error (MSE)

The **mean square error** of estimator $\hat{R}$ of parameter r is

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

This is an important measure of the *accuracy* of a point estimate.

Note that if $\hat{R}$ is an *unbiased estimate* of r then $\mathbf{E}[\hat{R}] = r$, which makes the MSE equal to $\mathbf{Var}[\hat{R}]$

Summary

If a sequence of *unbiased estimates* $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has *mean square error* $e_n = \mathbf{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then sequence $\hat{R}_n$ is **consistent**.

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

And from Chebyshev: $\lim_{n\to\infty} P(|\hat{R}_n - r| \geq \epsilon) \leq \lim_{n\to\infty} \frac{\mathbf{Var}[\hat{R}_n]}{\epsilon^2} = 0$,

therefore if variance strives to 0, then $\hat{R}_n$ is consistent.

Point Estimates of the Expected Value

Now, to estimate the parameter $r = \mathbf{E}[X]$ (the expected value of X) we use $\hat{R}_n = M_n(X)$, the sample mean.
For this we known that the sample mean $M_n(X)$ is an *unbiased estimate* of $\mathbf{E}[X]$

Then $M_n(X)$ has *mean square error*:

$$e_n = \mathbf{E}[(M_n(X) - \mathbf{E}[X])^2] = \mathbf{Var}[M_n(X)] = \frac{\mathbf{Var}[X]}{n}$$

(Here $\sqrt{e_n}$, is referred to as the **standard error** of the estimate, which is how far the sample mean will deviate from the expected value)
If X has finite variance, then the sample mean $M_n(X)$ is a sequence of **consistent estimates** of $\mathbf{E}[X]$.

Example 10.6

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $\mathrm{P}[A]$, has standard error ≤ 0.1?

Recap: Sample Mean

We can consider sampling as n trials, producing n iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean of X is the RV:

$$M_n(X) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\mathbf{E}[M_n(X)] = \mathbf{E}[X]$$

$$\mathbf{Var}[M_n(X)] = \frac{\mathbf{Var}[\mathrm{X}]}{n}$$

Note that this last one is NOT the sample variance, it is the variance of the sample mean.

Sample Variance

The **sample variance** of n independent observations of random variable X (with unknown μ_X) is

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

Sample Variance

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

Note that, unlike the sample mean, the sample variance is a *biased* estimate of $\mathbf{Var}[X]$ since:

$$\mathbf{E}[V_n(X)] = \frac{n-1}{n}\mathbf{Var}[X]$$

However, $V_n(X)$ is *asymptomatically unbiased* since:

$$\lim_{n\to\infty} \mathbf{E}[V_n(X)] = \lim_{n\to\infty} \frac{n-1}{n}\mathbf{Var}[X] = \mathbf{Var}[X]$$

Sample Variance

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

From the previous slide, we can "correct for the bias" to produce an *unbiased estimate* of $\mathbf{Var}[X]$, without the need for $n \to \infty$, hence $\mathbf{E}[V_n'(X)] = \mathbf{Var}[X]$

$$V_n'(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M_n(X))^2$$

Sometimes the top equation is called the **population variance**, and the bottom one is called the **sample variance**