AM216 Stochastic Differential Equations

Lecture 02

List of topics in this lecture

- Variance, properties of variance & expectation
- Bernoulli distribution, binomial distribution, normal distribution
- Memoryless process, derivation of exponential distribution
- Error function, calculation of confidence interval
- Interpretation of confidence interval

Review of probability theory (continued)

Variance:

$$\mathrm{var}(X) = E\left((X - E(X))^2\right) = E\left(X^2 - 2X\,E(X) + (E(X))^2\right)$$

$$= E(X^2) - 2E(X)E(X) + \left(E(X)\right)^2 = E(X^2) - \left(E(X)\right)^2$$

We obtain:

$$\boxed{\mathrm{var}(X) = E(X^2) - \left(E(X)\right)^2}$$

Standard deviation:

$$std(X) = \sqrt{\mathrm{var}(X)}$$

Properties of E(X) and var(X)

i) $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$

This is valid for all X and Y.

In particular, X and Y do not need to be independent.

ii) If X and Y are independent, then we have

$\mathrm{E}(X\,Y) = \mathrm{E}(X)\,\mathrm{E}(Y)$

Proof:

Independence implies

$$\rho_{(X,Y)}(x,y)=\rho_X(x)\rho_Y(y)$$

Using in the calculation of $E(XY)$, we get

$$E(XY)=\int xy\rho_{(X,Y)}(x,y)dx\,dy=\int xy\rho_X(x)\rho_Y(y)dx\,dy$$

$$=\left(\int x\rho_X(x)dx\right)\left(\int y\rho_Y(y)dy\right)=E(X)E(Y)$$

Caution:

- E(X Y) = E(X) E(Y) may not be true if X and Y are not independent.

  Example:

  $$X=Y=\begin{cases}2\,, & \Pr=0.5\\ 0\,, & \Pr=0.5\end{cases}$$

  E(X) = E(Y) = 2×0.5 = 1, E(X Y) = 4×0.5 = 2

  ==> E(X Y) ≠ E(X) E(Y)

- E(X Y) = E(X) E(Y) does not imply that X and Y are independent.

  Example:

  $$(X,Y)=\begin{cases}(0,1)\,, & \Pr=0.25\\ (0,-1)\,, & \Pr=0.25\\ (1,0)\,, & \Pr=0.25\\ (-1,0)\,, & \Pr=0.25\end{cases}$$

  E(X) = 0, E(Y) = 0, E(X Y) = 0

  ==> E(X Y) = E(X) E(Y)

  But $Y^2 = 1 - X^2$. So X and Y are definitely not independent of each other.

iii) If X and Y are independent, then we have

var(X + Y) = var(X) + var(Y)

Proof:

var(X + Y) = E((X + Y)^2) − (E (X + Y)) ^{2} = ⋯

Complete the proof in your homework.

Examples of distributions:

1) Bernoulli distribution

Consider the number of success in ONE trial with success probability p

$$X = \begin{cases} 1, & \Pr = p \\ 0, & \Pr = 1-p \end{cases}$$

Random variable X has the Bernoulli distribution with parameter p.

Notation:

$X \sim \text{Bern}(p)$

Range = {0, 1}.

Example: Flip a coin

1: head, success

0: tail, failure

Expected value and variance:

$E(X) = p$, $E(X^2) = p$

$\text{Var}(X) = E(X^2) - (E(X))^2 = p(1-p)$

2) Binomial distribution

Consider the number of successes in a sequence of n independent trials, each with success probability p.

N = sum of n independent Bernoulli random variables

$$N = \sum_{i=1}^{n} X_i$$

Random variable N has the binomial distribution with parameters (n, p).

Notation:

$N \sim \text{Bino}(n, p)$ or simply $N \sim B(n, p)$

Range = $\{0, 1, 2, \ldots, n\}$.

PMF (probability mass function):

$$\Pr(N = k) = C(n,k) p^k (1-p)^{n-k}$$

Example: # of heads in n flips of a coin

Expected value and variance:

$$E(N) = E\left(X_1 + X_2 + \cdots + X_n\right) = np$$

$$\text{var}(N) = \text{var}\left(X_1 + X_2 + \cdots + X_n\right) = np\left(1-p\right)$$

3) Exponential distribution

Example: (Escape problem)

T = time until escape from a deep potential well by thermal fluctuations

PDF (probability density function):

$$\rho_T(t) = \begin{cases} \lambda \exp\left(-\lambda t\right), & t \ge 0 \\ 0, & t < 0 \end{cases}$$

Random variable T has the exponential distribution with parameter λ.

Notation:

$T \sim \text{Exp}(\lambda)$

Range = $(0, +\infty)$.

Mathematical definition of exponential distribution:

T = time from $t = 0$ until occurrence of an event in a memoryless system

Derivation of PDF for exponential distribution

We derive the PDF based on the "memoryless" property.

Recall that T = time until occurrence. "Memoryless" means

"Given that the event has not occurred at t_0, the additional time until occurrence is not affected by t_0 no matter how large or how small t_0 is."

==> $\Pr\left((T-t_0) \le t \middle| T > t_0\right) = \Pr\left(T \le t\right)$

Consider the complementary cumulative distribution function (CCDF)

$$G(t) \equiv \Pr\left(T > t\right) = \int_t^\infty \rho(t')dt'$$

$G(0) = \Pr(T > 0) = 1$

We re-write the memoryless property in terms of $G(t)$.

$$\frac{\Pr\left((T-t_0)\le t \text{ AND } T>t_0\right)}{\Pr\left(T>t_0\right)}=\Pr\left(T\le t\right)$$

$$\Longrightarrow \quad \Pr\left(t_0<T\le t_0+t\right)=\Pr\left(T\le t\right)\Pr\left(T>t_0\right)$$

$$\Longrightarrow \quad G(t_0)-G(t_0+t)=\left(1-G(t)\right)G(t_0)$$

Replace t with Δt, divide by Δt, and take the limit as $\Delta t\to 0$, we get

$$\frac{G(t_0)-G(t_0+\Delta t)}{\Delta t}=\frac{G(0)-G(\Delta t)}{\Delta t}G(t_0)$$

$$\Longrightarrow \quad G'(t_0)=\underbrace{G'(0)}_{-\lambda}G(t_0)$$

We obtain an initial value problem (IVP) for $G(t_0)$

$$\begin{cases} G'(t_0)=-\lambda G(t_0) \\ G(0)=1 \end{cases}$$

The solution is $\quad G(t)=\exp(-\lambda t)$.

Differentiate $G(t)\equiv\int_t^\infty \rho(t')dt'$, we obtain

$$\rho(t)=-\frac{d}{dt}G(t)=\begin{cases} \lambda\exp(-\lambda t), & t\ge 0 \\ 0, & t<0 \end{cases}$$

<u>Expected value and variance:</u>

$$E(T)=\frac{1}{\lambda}, \qquad \mathrm{var}(T)=\frac{1}{\lambda^2}$$

<u>CDF:</u>

$$F_T(t)=\Pr\left(T\le t\right)=1-\exp\left(-\lambda t\right) \qquad \text{for } t\ge 0$$

<u>4) Normal distribution</u>

<u>PDF:</u>

$$\rho_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

Random variable X has the normal distribution with parameters (μ, σ^2).

Notation:

$X \sim N(\mu, \sigma^2)$

Range = $(-\infty, +\infty)$

Example: (*Central Limit Theorem*)

Suppose $\{X_1, X_2, \ldots, X_M\}$ are i.i.d. (independent and identically distributed).

When M is large, $X = \sum_{j=1}^{M} X_j$ approximately has a normal distribution.

Expected value and variance:

$$E(X) = \int x\rho(x)dx = \mu$$

$$\text{var}(X) = E\left((X-\mu)^2\right) = \int (x-\mu)^2 \rho(x)dx = \sigma^2$$

CDF of normal distribution:

$$F_X(x) = \Pr\left(X \le x\right) = \int_{-\infty}^{x} \rho_X(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x'-\mu)^2}{2\sigma^2}\right)dx'$$

Change of variables: $\quad s = \dfrac{x'-\mu}{\sqrt{2\sigma^2}}, \quad dx' = \sqrt{2\sigma^2}\,ds$

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sqrt{2\sigma^2}}} \frac{1}{\sqrt{\pi}} \exp(-s^2)ds = \frac{1}{2} + \int_{0}^{\frac{x-\mu}{\sqrt{2\sigma^2}}} \frac{1}{\sqrt{\pi}} \exp(-s^2)ds$$

We write the CDF in terms of the error function.

The error function:

$$\text{erf}(z) \equiv \frac{1}{\sqrt{\pi}} \int_{-z}^{z} \exp(-s^2)ds = \frac{2}{\sqrt{\pi}} \int_{0}^{z} \exp(-s^2)ds$$

Properties of erf(z):

i) $\text{erf}(0) = 0$

ii) $\text{erf}(+\infty) = 1$

iii) $\text{erf}(-z) = -\text{erf}(z)$

The CDF of normal distribution has the expression

$$\boxed{F_X(x) = \frac{1}{2}\left(1+\text{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right)}$$

Example:

$$\Pr\left(X \le \mu+\eta\sigma\right) = \frac{1}{2}\left(1+\text{erf}\left(\frac{\mu+\eta\sigma-\mu}{\sqrt{2\sigma^2}}\right)\right) = \frac{1}{2}\left(1+\text{erf}\left(\frac{\eta}{\sqrt{2}}\right)\right)$$

We like to find η such that

$$\Pr\left(\left|X-\mu\right| \le \eta\sigma\right) = 0.95 \qquad (95\%)$$

We express this probability in terms of CDF, and then in terms of erf( ).

$$\Pr\left(\left|X-\mu\right| \le \eta\sigma\right) = \Pr\left(\mu-\eta\sigma \le X \le \mu+\eta\sigma\right)$$

$$= F_X\left(\mu+\eta\sigma\right) - F_X\left(\mu-\eta\sigma\right) = \quad \cdots \quad = \text{erf}\left(\frac{\eta}{\sqrt{2}}\right)$$

Setting $\text{erf}\left(\frac{\eta}{\sqrt{2}}\right) = 0.95$, we calculate η as

$$\eta = \text{erfinv}\left(0.95\right)\sqrt{2} = 1.96$$

We obtain

$$\boxed{\Pr\left(\left|X-\mu\right| \le 1.96\sigma\right) = 95\%}$$

Similarly, we can obtain

$$\boxed{\Pr\left(\left|X-\mu\right| \le 2.5758\sigma\right) = 99\%}$$

Confidence interval:

Suppose we are given a data set of n independent samples of $X \sim N(\mu, \sigma^2)$.

$\{X_j, j = 1, 2, \ldots, n\}$

Suppose we don't know μ.

Question: How to estimate μ from data?

We can use the sample mean.

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

<u>Question:</u> How to estimate the uncertainty/error in $\hat{\mu}$?

$\hat{\mu}$ is a random variable, derived from random variables (X_1, X_2, ..., X_n).

$$E(\hat{\mu}) = E\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n}E\left(X_1 + \cdots + X_n\right) = \mu$$

$$\mathrm{var}(\hat{\mu}) = \mathrm{var}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = \frac{1}{n^2}\mathrm{var}\left(X_1 + \cdots + X_n\right) = \frac{\sigma^2}{n}$$

(Here we used the independence of $\{X_j\}$)

<u>Theorem:</u>

Sum of independent normal random variables is a normal random variable

This theorem will be proved in the discussion of characteristic functions.

It follows from this theorem that $\hat{\mu}$ is normal.

$$\hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

The interval containing 95% probability of $\hat{\mu}$ is described by

$$\Pr\left(\left|\hat{\mu} - \mu\right| \le 1.96\frac{\sigma}{\sqrt{n}}\right) = 95\%$$

<u>Case 1:</u> Suppose we know the value of σ.

$\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}},\ \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ is called the 95% confidence interval.

<u>Example:</u>

We are given a data set of 100 independent samples of $X \sim N(\mu, \sigma^2)$:

{3.0811, 0.7589, 1.9611, ... }

We are given σ = 1.3.

μ is estimated as

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j = 0.475$$

$$1.96\frac{\sigma}{\sqrt{n}} = 0.2548$$

The 95% confidence interval is (0.2202, 0.7298)

Interpretation of the confidence interval

Question: What is the meaning of this confidence interval?

The data set is given, fixed.

μ is fixed, although unknown.

What is the meaning of interval (0.2202, 0.7298)?

Two key components in interpreting the confidence interval:

i) The confidence interval is an algorithm/function that maps a data set $\{X_j\}$ to an interval

$$\{X_j\} \longrightarrow \left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$$

It is important to notice that interval $\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$ varies with data set.

If we view the data as a set of random samples, than the interval $\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$ is a random variable, derived from the random samples.

ii) The framework of repeated experiments.

Draw a data set of n independent samples of $X \sim N(\mu^{(\text{True})}, \sigma^2)$.

Repeat the drawing M times (M is large).

The meaning of confidence interval is

$$\Pr\left(\underbrace{\hat{\mu}_L(\{X_j\})}_{\substack{\text{Random}\\ \text{variable}}} < \underbrace{\mu^{(\text{True})}}_{\text{Fixed}} < \underbrace{\hat{\mu}_H(\{X_j\})}_{\substack{\text{Random}\\ \text{variable}}} \right) = 0.95$$

When we go over M data sets and estimate the confidence interval for each data set, for 95% of data sets, the estimated confidence interval contains $\mu^{(\text{True})}$.

In summary, the two key components for interpreting the confidence interval are

i) it is an algorithm mapping a data set to an interval,

ii) the framework of repeated experiments

Case 2: σ is unknown

Recall the definition of standard deviation.

$$\sigma = \sqrt{\text{var}(X)} = \sqrt{E\left((X-\mu)^2\right)}$$

From the given samples, we can calculate the sample standard deviation

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\mu})^2}\ , \qquad \hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

<u>Note:</u> The denominator is $(n-1)$ instead of n.

This modification is to make the sample standard deviation unbiased.

We write an <u>approximate</u> 95% confidence interval is

$$\left(\hat{\mu} - 1.96\frac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

Correspondingly, the one calculated using the exact value of σ (case 1 above) is called the <u>exact</u> 95% confidence interval

$$\left(\hat{\mu} - 1.96\frac{\sigma}{\sqrt{n}},\ \hat{\mu} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$