**AM216 Stochastic Differential Equations**

Lecture 02


**List of topics in this lecture**

- Variance, properties of expectation and variance
- Bernoulli distribution, binomial distribution, memoryless process, derivation of PDF of exponential distribution, normal distribution
- CDF of normal distribution, error function, confidence interval for the mean
- Interpretation of confidence interval

---

## Recap

The framework of repeated experiments for probability: without specifying how the experiment is repeated, the concept of probability does not make sense.

Properties of expected value:

$E(\alpha X) = \alpha E(X)$

$E(X+Y) = E(X) + E(Y)$ for all $X$ and $Y$

Law of total probability: $\Pr(A) = \sum_n \Pr(A|B_n)\Pr(B_n)$

Law of total expectation: $E(X) = E\Big(E(X|Y)\Big)$, $E(X) = \sum_n E(X|B_n)\Pr(B_n)$

## Review of probability theory (continued)

<u>Variance:</u>

$$\text{var}(X) \equiv E\Big((X-E(X))^2\Big) = E\Big(X^2 - 2E(X)X + (E(X))^2\Big)$$

Recall that $E(X)$ is a deterministic number

$$= E(X^2) - 2E(X)E(X) + \Big(E(X)\Big)^2 = E(X^2) - \Big(E(X)\Big)^2$$

We obtain:

$$\boxed{\text{var}(X) = E(X^2) - \Big(E(X)\Big)^2}$$

Standard deviation:

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

Properties of $E(X)$

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 the independence 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:

- $\mathrm{E}(X\,Y) = \mathrm{E}(X)\,\mathrm{E}(Y)$ may not be true if $X$ and $Y$ are not independent.

  Example:

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

  We have $\mathrm{E}(X) = \mathrm{E}(Y) = 2\times 0.5 = 1$, $\quad \mathrm{E}(X\,Y) = 4\times 0.5 = 2$

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

- $\mathrm{E}(X\,Y) = \mathrm{E}(X)\,\mathrm{E}(Y)$ does not imply that $X$ and $Y$ are independent.

  Example:

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

  We have $\mathrm{E}(X) = 0$, $\quad \mathrm{E}(Y) = 0$, $\quad \mathrm{E}(X\,Y) = 0$

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

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

Properties of var($X$)

iii) $\mathrm{var}(\alpha X) = \alpha^2 \mathrm{var}(X)$

Proof is in your homework.

iv) If $X$ and $Y$ are independent, then we have

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$$

Proof:

$$\mathrm{var}(X+Y) = \mathrm{E}((X+Y)^2) - (\mathrm{E}(X+Y))^2 = \cdots$$

Complete the proof in your homework.

Examples of distributions:

1) Bernoulli distribution

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

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

We say random variable $X$ has the Bernoulli distribution with parameter $p$.

Notation:

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

Range = {0, 1}.

Example: Flip a coin

1: head, success

0: tail, failure

Expected value and variance:

$$\mathrm{E}(X) = 0\times(1-p)+1\times p = p, \qquad \mathrm{E}(X^2) = p$$

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

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 with parameter $p$.

$$N = \sum_{i=1}^{n} X_i \,, \qquad X_i \sim \text{(iid) Bern}(p)$$

iid = independently and identically distributed

We say random variable $N$ has the binomial distribution with parameters $(n, p)$.

Notation:

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

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

PMF (probability mass function):

$$\Pr(N = k) = C(n,k) p^k (1-p)^{n-k} \,, \qquad k = 0, 1, 2, \ldots, n$$

Example: # of heads in $n$ flips of a coin

Expected value and variance:

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

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

3) Exponential distribution

Example: (Escape problem)

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

PDF of $T$ has the form

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

We say random variable $T$ has the exponential distribution with parameter $\lambda$.

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 of $T$ 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."

==> $\Pr\big(\underbrace{(T-t_0)}_{\substack{\text{additional}\\ \text{time}}}\le t \,\big|\, T>t_0\big)=\Pr(T\le t)$

Consider the complementary cumulative distribution function (CCDF)

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

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

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

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

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

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

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

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

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

Let $\lambda \equiv -G'(0)$. We obtain an initial value problem (IVP) for $G(t_0)$

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

The solution is $G(t)=\exp(-\lambda t),\quad t>0$.

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

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

<u>Remark:</u> The time until occurrence of an event in a memoryless system must have a PDF of the form given above.

Expected value and variance:

$$E(T) = \int t\rho_T(t)dt = \int_0^{+\infty} t\lambda \exp(-\lambda t)dt = \frac{1}{\lambda}$$

$$E(T^2) = \int_0^{+\infty} t^2 \lambda \exp(-\lambda t)dt = \frac{2}{\lambda^2} \qquad \text{(see Appendix A for the calculation)}$$

$$\text{var}(T) = E(T^2) - E(T)^2 = \frac{1}{\lambda^2}$$

CDF:

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

4) Normal distribution

PDF:

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

We say random variable $X$ has the normal distribution with parameters $(\mu, \sigma^2)$.

Notation:

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

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

Example: (*Central Limit Theorem*)

Suppose $\{X_1, X_2, \ldots, X_M\}$ are iid (independently and identically distributed).

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

Expected value and variance:

$$E(X) = E(X-\mu) + \mu = \underbrace{\int (x-\mu)\rho_X(x)dx}_{=0 \text{ because of symmetry}} + \mu = \mu$$

$$\text{var}(X) = E\left((X-\mu)^2\right) = \int (x-\mu)^2 \rho_X(x)dx = \sigma^2 \qquad \text{(see Appendix A)}$$

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: $s = \dfrac{x'-\mu}{\sqrt{2\sigma^2}}, \qquad dx' = \sqrt{2\sigma^2}\, ds$

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

We write the CDF in terms of the error function.

<u>The error function:</u>

$$\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z \exp(-s^2) ds$$

<u>Properties of erf(z):</u>

i) erf(0) = 0

ii) erf(+∞) = 1

iii) erf(−z) = −erf(z)

In terms of erf($z$), the CDF of normal distribution has the expression

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

<u>Example:</u>

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

<u>Interval containing 95% probability</u>

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( ).

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

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

Setting $\mathrm{erf}\left(\frac{\eta}{\sqrt{2}}\right)=0.95$, we calculate η using the inverse error function

$$\eta=\mathrm{erfinv}(0.95)\sqrt{2}=1.96$$

We obtain

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

Similarly, we can obtain

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

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 and we want to estimate μ from the data.

Question: How to estimate μ from data?

We can use the sample mean.

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

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

First we recognize that $\hat{\mu}$ is a random variable, derived from random variables ($X_1$, $X_2$, …, $X_n$). Each data set gives a (potentially) different value of $\hat{\mu}$.

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

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

Here we used the independence of $\{X_j\}$.

Theorem:

Sum of independent normal random variables is a normal random variable.

Proof: It will be proved in the discussion of characteristic functions.

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

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

$$\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$ This is called a standard normal.

The interval containing 95% probability is

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

<u>Case 1:</u> Suppose the value of σ is given.

$$\left|\frac{\hat{\mu}-\mu}{\sigma/\sqrt{n}}\right| \le 1.96 \qquad \Longleftrightarrow \qquad \mu \in \left(\hat{\mu}-1.96\frac{\sigma}{\sqrt{n}},\ \hat{\mu}+1.96\frac{\sigma}{\sqrt{n}}\right)$$

which is called the 95% confidence interval (CI) for the mean.

<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, 0.3050, 0.3887, 1.4971, 1.3225, −0.8563, … }

We are given σ = 1.3. We estimate μ using the sample mean

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

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

The 95% confidence interval for the mean is (0.2202, 0.7298)

<u>Interpretation of the confidence interval</u>

<u>Question:</u> What is the meaning of the 95% confidence interval for the mean?

μ is fixed, although unknown. μ is not random.

For the given data set, the 95% confidence interval is determined: (0.2202, 0.7298).

We have either μ ∈ (0.2202, 0.7298) or μ ∉ (0.2202, 0.7298).

It is not uncertain. It is just unknown to us (because μ is unknown).

"Pr(μ ∈ (0.2202, 0.7298)) = 95%" does not make sense.

<u>Two key components in interpreting the confidence interval:</u>

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

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

where $\quad \hat{\mu}_L(\{X_j\}) = \hat{\mu}(\{X_j\}) - 1.96\dfrac{\sigma}{\sqrt{n}},\quad \hat{\mu}_H(\{X_j\}) = \hat{\mu}(\{X_j\}) + 1.96\dfrac{\sigma}{\sqrt{n}}$

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

For a random data set, $\left(\hat{\mu}_L(\{X_j\}),\ \hat{\mu}_H(\{X_j\})\right)$ is a random variable, derived from the random data set.

ii) We view it in the framework of repeated experiments.

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

Repeat the drawing $M$ times ($M$ is large).

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 μ.

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

<u>In summary,</u> the two key components for interpreting the confidence interval are

i) the confidence interval is <u>an algorithm</u> mapping a data set to an interval; and

ii) the 95% probability is in the framework of hypothetically drawing a large number of data sets and applying the algorithm to each data set.

<u>Case 2:</u> σ is unknown

Recall the definition of standard deviation.

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

From the given samples, we can calculate the <u>sample standard deviation</u>

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

<u>Note:</u> The denominator is $(n-1)$ instead of $n$. This modification is to make the sample variance unbiased: $E(\hat{\sigma}^2) = \sigma^2$.

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n-1}\sum_{j=1}^{n}(X_j - \hat{\mu})^2\right) = \frac{1}{(n-1)}\sum_{j=1}^{n} E\left((X_j - \hat{\mu})^2\right), \qquad \hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

Let $Y_j \equiv X_j - \mu$. We have

$$X_j = \mu + Y_j\,,\quad \hat{\mu} = \mu + \frac{1}{n}\sum_{k=1}^{n} Y_k\,,\quad E(Y_k) = 0 \text{ and } E(Y_k^{\,2}) = \sigma^2$$

$$E\Big((X_1 - \hat{\mu})^2\Big) = E\left((Y_1 - \frac{1}{n}\sum_{k=1}^{n} Y_k)^2\right) = E\left(\left(\frac{n-1}{n}Y_1 - \frac{1}{n}Y_2 - \cdots - \frac{1}{n}Y_n\right)^2\right)$$

$$= E\left(\frac{(n-1)^2}{n^2}Y_1^{\,2} + \frac{1}{n^2}Y_2^{\,2} + \cdots + \frac{1}{n^2}Y_n^{\,2}\right) = \left(\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right)\sigma^2 = \frac{n-1}{n}\sigma^2$$

$$E(\hat{\sigma}^2) = \frac{1}{(n-1)}\sum_{j=1}^{n} E\Big((X_j - \hat{\mu})^2\Big) = \frac{1}{(n-1)} n \frac{(n-1)}{n}\sigma^2 = \sigma^2$$

Using $\hat{\sigma}$, we write out an approximate 95% confidence interval

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

A better solution for case 2 (optional):

When $\sigma$ is unknown, we use $\hat{\sigma}$ to replace $\sigma$. $\dfrac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{n}}$ is not exactly a normal distribution (it is approximately a normal distribution).

$\dfrac{\hat{\mu} - \mu}{(\hat{\sigma}/\sqrt{n})}$ is exactly a Student's $t$-distribution with $(n-1)$ degrees of freedom.

From the inverse CDF of the $t$-distribution, we can find the exact value of $\eta$ such that

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

$$\Longleftrightarrow \quad F_t\big(\eta, (n-1)\big) = 0.975$$

$$\Longleftrightarrow \quad \eta = F_t^{(\text{inv})}\big(0.975, (n-1)\big)$$

The 95% confidence interval is $\left(\hat{\mu} - \eta\dfrac{\hat{\sigma}}{\sqrt{n}},\ \hat{\mu} + \eta\dfrac{\hat{\sigma}}{\sqrt{n}}\right)$.

**Appendix A:** An alternative way of calculating some integrals

Integral 1: $I_1 = \int_0^{+\infty} t^2 \lambda \exp(-\lambda t) dt$

To calculate $I_1$, we consider

$$G(\lambda) \equiv \int_0^{+\infty} \exp(-\lambda t) dt = \frac{1}{\lambda}, \qquad \frac{dG(\lambda)}{d\lambda} = -\int_0^{+\infty} t \exp(-\lambda t) dt = \frac{-1}{\lambda^2}$$

We write $I_1$ as

$$I_1 = \lambda \int_0^{+\infty} t^2 \exp(-\lambda t) dt = \lambda \frac{d^2 G(\lambda)}{d\lambda^2} = \lambda \frac{2}{\lambda^3} = \frac{2}{\lambda^2}$$

Integral 2: $I_2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx$

To calculate $I_2$, we consider

$$G(\sigma) \equiv \int_{-\infty}^{+\infty} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \sqrt{2\pi\sigma^2}, \qquad \frac{dG(\sigma)}{d\sigma} = \frac{1}{\sigma^3} \int_{-\infty}^{+\infty} x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \sqrt{2\pi}$$

We write $I_2$ as

$$I_2 = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \frac{\sigma^2}{\sqrt{2\pi}} \frac{dG(\sigma)}{d\sigma} = \sigma^2$$