Systems Engineering 520

Review of Statistics and Probability

Huseyin Topaloglu
School of Operations Research
and Industrial Engineering

Lecture 2

Cornell University

Overview

- Discrete and continuous random variables
- Expectation and variance
- Important discrete random variables
- Important continuous random variables
- Central limit theorem
- Confidence intervals
- Gradient of a function

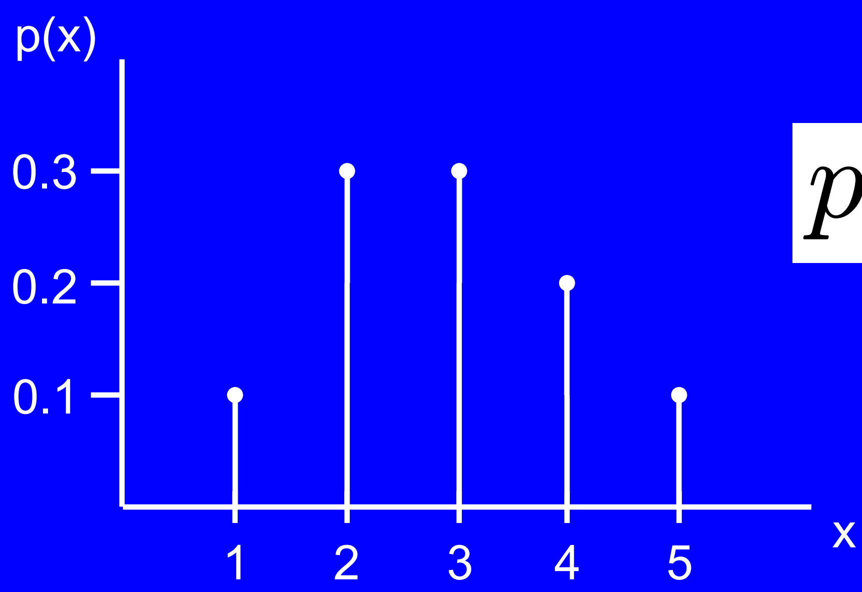
Random Variables

- A random variable is a function whose value is determined by the outcome of a random experiment
- More technically, a random variable is a function from the set of outcomes of a random experiment to the real numbers
- E.g. for the random experiment of rolling a die, define the random variable X as

$$X = \begin{cases} 1 & \text{if we get an even number} \\ 0 & \text{if we get an odd number} \end{cases}$$

Discrete Random Variables

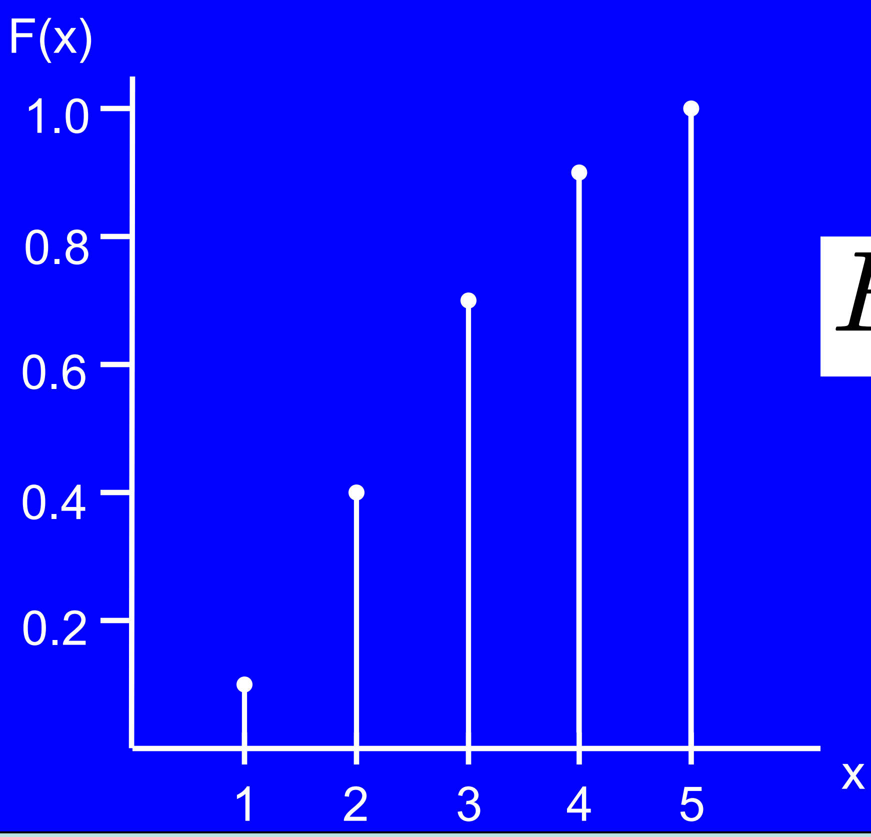
- Discrete random variables take on countably many (discrete) values
- A discrete random variable X is characterized by its probability mass function p(.)



$$p(x) = \mathbb{P}\{X = x\}$$

Discrete Random Variables

- A discrete random variable X can also be characterized by its cumulative distribution function F(.)



$$F(x) = \mathbb{P}\{X \leq x\}$$

Discrete Random Variables

- Knowledge of p(.) and knowledge of F(.) are equivalent to each other

$$
\begin{aligned}
F(x) &= \mathbb{P}\{X \leq x\} \\
&= \sum_{i=-\infty}^{x} \mathbb{P}\{X = i\} = \sum_{i=-\infty}^{x} p(i)
\end{aligned}
$$

$$
\begin{aligned}
p(x) &= \sum_{i=-\infty}^{x} p(i) - \sum_{i=-\infty}^{x-1} p(i) \\
&= \mathbb{P}\{X \leq x\} - \mathbb{P}\{X \leq x-1\} \\
&= F(x) - F(x-1)
\end{aligned}
$$

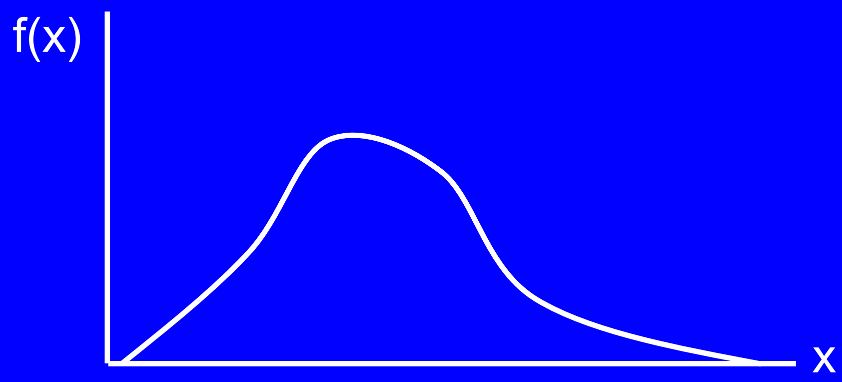
Discrete Random Variables

- Also note that

$$\begin{aligned}\mathbb{P}\{a \le X \le b\} &= \sum_{i=a}^{b} p(i) \\ &= \sum_{i=-\infty}^{b} p(i) - \sum_{i=-\infty}^{a-1} p(i) \\ &= F(b) - F(a-1)\end{aligned}$$

Continuous Random Variables

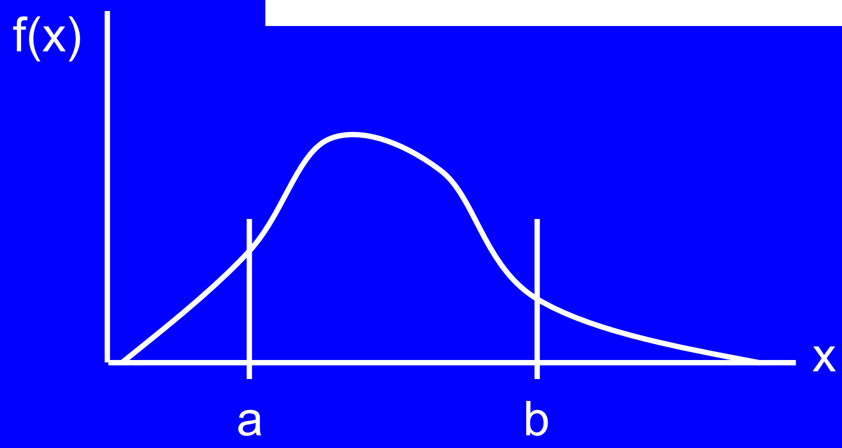
- Continuous random variables take on values on a continuum
- E.g. closing price of a stock next day, lifetime of a light bulb
- A continuous random variable X is characterized by its probability density function f(.)



Continuous Random Variables

- It is not true that $f(x) = \mathbb{P}\{X = x\}$
- In fact, for any value x, $\mathbb{P}\{X = x\} = 0$
- But we have

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b f(x)\,dx$$



Continuous Random Variables

- The cumulative distribution function of a random variable X is defined as

$$F(x) = \mathbb{P}\{X \leq x\} = \mathbb{P}\{-\infty \leq X \leq x\}$$
$$= \int_{\infty}^{x} f(u)\, du$$

- If f(.) is continuous in an interval containing x,

$$f(x) = F'(x)$$

Continuous Random Variables

- Also note that

$$\begin{aligned}\mathbb{P}\{a \leq X \leq b\} &= \int_a^b f(u)\,du \\ &= \int_{-\infty}^b f(u)\,du - \int_{-\infty}^a f(u)\,du \\ &= F(b) - F(a)\end{aligned}$$

Expectation

- If X is a discrete random variable with p.m.f. p(.)

$$\mathbb{E}\{X\} = \sum_{x=-\infty}^{\infty} x\, p(x)$$

- If X is a continuous random variable with p.d.f. f(.)

$$\mathbb{E}\{X\} = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Expectation

- For any two random variables X and Y,

$$\mathbb{E}\{a\,X\} = \int_{-\infty}^{\infty} a\,x\,f(x)\,dx$$

$$a\int_{-\infty}^{\infty} x\,f(x)\,dx = a\,\mathbb{E}\{X\}$$

$$\mathbb{E}\{a\,X + b\,Y\} = a\,\mathbb{E}\{X\} + b\,\mathbb{E}\{Y\}$$

Linearity of expectation

Variance

- The variance of a random variable X is

$$Var(X) = \mathbb{E}\{[X - \mathbb{E}\{X\}]^2\}$$

- Also note that

$$\begin{aligned}
Var(X) &= \mathbb{E}\{[X - \mathbb{E}\{X\}]^2\} \\
&= \mathbb{E}\left\{X^2 - 2\,X\,\mathbb{E}\{X\} + [\mathbb{E}\{X\}]^2\right\} \\
&= \mathbb{E}\{X^2\} - 2\,\mathbb{E}\left\{X\,\mathbb{E}\{X\}\right\} + [\mathbb{E}\{X\}]^2 \\
&= \mathbb{E}\{X^2\} - 2\,\mathbb{E}\{X\}\,\mathbb{E}\{X\} + [\mathbb{E}\{X\}]^2 \\
&= \mathbb{E}\{X^2\} - [\mathbb{E}\{X\}]^2
\end{aligned}$$

Variance

- If X is a discrete random variable with p.m.f. p(.) and g(.) is a real valued function, then

$$\mathbb{E}\{g(X)\} = \sum_{x=-\infty}^{\infty} g(x)\, p(x)$$

- If X is a continuous random variable with p.d.f. f(.) and g(.) is a real valued function, then

$$\mathbb{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Variance

- Assume that X is a continuous random variable with p.d.f.

$$f(x) = \begin{cases} x^2/9 & \text{if } 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

- Compute c.d.f. of X, E{X}, E{X^2} and Var(X)

Variance

- For two independent random variables X and Y

$$Var(a\,X) = a^2\,Var(X)$$

$$Var(a\,X + b\,Y) = a^2\,Var(X) + b^2\,Var(Y)$$

$$\begin{aligned} Var(aX) &= \mathbb{E}\{[a\,X]^2\} - [\mathbb{E}\{a\,X\}]^2 \\ &= \mathbb{E}\{a^2\,X^2\} - [a\,\mathbb{E}\{X\}]^2 \\ &= a^2\ \{\mathbb{E}\{X^2\} - \mathbb{E}\{X\}^2\} \end{aligned}$$

Bernoulli Random Variable

- For some p in [0,1], if we have

$$\mathbb{P}\{X = 1\} = p \text{ and } \mathbb{P}\{X = 0\} = 1 - p$$

then X has Bernoulli distribution with parameter p

- If X has Bernoulli distribution with parameter p, then

$$\begin{aligned} \mathbb{E}\{X\} &= \\ Var(X) &= \end{aligned}$$

Geometric Random Variable

- For some p in [0,1], if we have

$$\mathbb{P}\{X = k\} = (1-p)^{k-1}\,p$$

then X has geometric distribution with parameter p

- X takes values 1, 2, 3,…

- If X has geometric distribution with parameter p, then

$$\mathbb{E}\{X\} = 1/p$$

$$Var(X) = [1-p]/p^2$$

Geometric Random Variable

- Imagine making successive independent trials
- Each trial is a success with probability p and is failure with probability 1-p
- Let X be the number of trials to get the first success

$$\mathbb{P}\{X = k\} = \underbrace{(1-p)\,(1-p)\,\ldots\,(1-p)}_{\text{k-1 times}}\,p$$

- So a geometric random variable can be visualized as the number of trials to get the first success

Uniform Random Variable

- For some interval [a,b], if a random variable X has the p.d.f.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

then X is has uniform distribution over [a,b]

- If X has uniform distribution over [a,b], then

$$\mathbb{E}\{X\} = [a+b]/2$$

$$Var(X) = [b-a]^2/12$$

Exponential Random Variable

- For some $\lambda > 0$, if a random variable X has the p.d.f.

$$f(x) = \lambda e^{-\lambda x}$$

then X has exponential distribution with parameter λ

- If X has exponential distribution with parameter λ, then

$$\mathbb{E}\{X\} = 1/\lambda$$

$$Var(X) = 1/\lambda^2$$

Exponential Random Variable

- If X has exponential distribution with parameter λ, then the c.d.f of X is

$$F(x) = \int_0^x \lambda\, e^{-\lambda\, u}\, du = 1 - e^{-\lambda\, x}$$

Normal Random Variable

- For some μ and $\sigma > 0$, if a random variable X has the p.d.f.

$$f(x) = \frac{1}{[2\,\pi]^{1/2}\,\sigma}\, e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2}$$

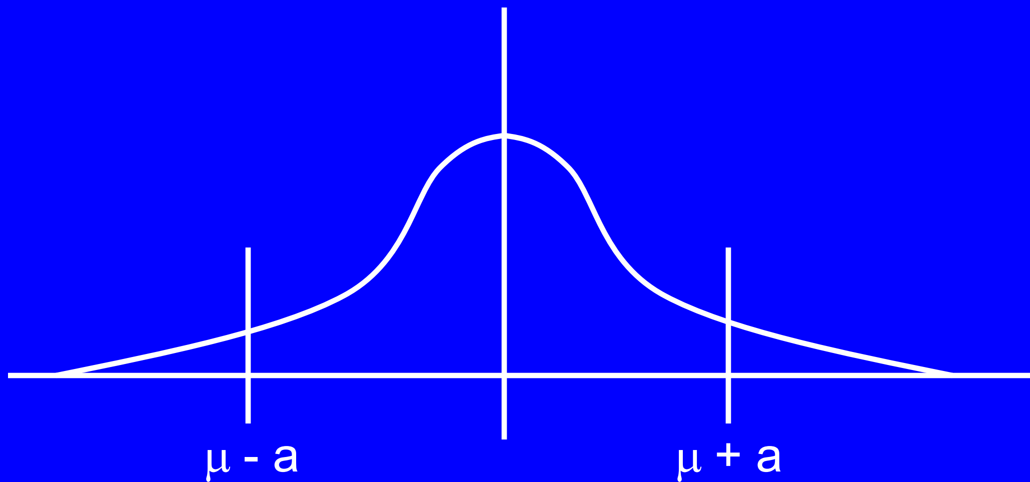
  then X has normal distribution with mean μ and standard deviation σ

- If X has normal distribution with mean μ and standard deviation σ, then

$$\mathbb{E}\{X\} = \mu$$
$$Var(X) = \sigma^2$$

Normal Random Variable

- The probability density function is symmetric around the mean μ



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

$$\mathbb{P}\{X \leq \mu - a\} = \mathbb{P}\{X \geq \mu + a\}$$

Normal Random Variable

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

$$aX + b \sim N(a\,\mu + b, a^2\,\sigma^2)$$

$$\frac{X-\mu}{\sigma} \sim \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

$$\sim N\left(\frac{1}{\sigma}\mu - \frac{\mu}{\sigma}, \frac{1}{\sigma^2}\sigma^2\right) \sim N(0,1)$$

- If $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, and X and Y are independent of each other, then

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

Normal Random Variable

- If X~N(μ,σ^2), then

$$\begin{aligned}\mathbb{P}\{X \le x\} &= \mathbb{P}\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} \\ &= \mathbb{P}\left\{N(0,1) \le \frac{x-\mu}{\sigma}\right\}\end{aligned}$$

- Therefore, if we know the cumulative distribution function for N(0,1), then we can compute the cumulative distribution function for any normal random variable

Central Limit Theorem

- Let X_1, X_2, X_3,… be independent random variables each having mean μ and variance σ^2
- Define S_n as $S_n = X_1 + \ldots + X_n$

$$
\begin{aligned}
\mathbb{E}\{S_n\} &= \mathbb{E}\{X_1 + \ldots + X_n\} \\
&= \mathbb{E}\{X_1\} + \ldots + \mathbb{E}\{X_n\} = n\,\mu
\end{aligned}
$$

$$
Var(S_n) = n\,\sigma^2
$$

- What does the distribution of S_n look like?
- Distribution of S_n should depend on the distributions of X_1, X_2, X_3,…

Central Limit Theorem

- Simulation results indicate that the distribution of S_n looks like the normal distribution as n becomes large
- This is the essence of the central limit theorem
- Let $X_1, X_2, X_3, \ldots$ be independent random variables each having mean μ and variance σ^2
- Define $S_n = X_1 + \ldots + X_n$
- When n is large, the distribution of S_n is approximately normal

$$S_n \approx N(n\,\mu, n\,\sigma^2)$$

Building a Confidence Interval for the Expected Value of a Random Variable

- Central limit theorem can be used to build confidence intervals for the expected value of a random variable
- Let X_1, X_2, X_3,… be independent random variables each having mean μ and variance σ^2
- Assume that we do not know μ and want to estimate it
- Define

$$\bar{X}_n = \frac{1}{n}[X_1 + \ldots + X_n] = \frac{S_n}{n}$$

Building a Confidence Interval for the Expected Value of a Random Variable

- When n is large

$$\bar{X}_n = \frac{S_n}{n} \approx N\left(\frac{n\,\mu}{n}, \frac{n\,\sigma^2}{n^2}\right) = N(\mu, \sigma^2/n)$$

- From the cumulative distribution of standard normal random variable, we have

$$\mathbb{P}\{-1.96 \leq N(0,1) \leq 1.96\} = 0.95$$

Building a Confidence Interval for the Expected Value of a Random Variable

$$\mathbb{P}\{-1.96 \leq N(0,1) \leq 1.96\} = 0.95$$

$$\mathbb{P}\left\{-1.96\frac{\sigma}{n^{1/2}} \leq \frac{\sigma}{n^{1/2}}N(0,1) \leq 1.96\frac{\sigma}{n^{1/2}}\right\} = 0.95$$

$$\mathbb{P}\left\{\mu - 1.96\frac{\sigma}{n^{1/2}} \leq \mu + \frac{\sigma}{n^{1/2}}N(0,1) \leq \mu + 1.96\frac{\sigma}{n^{1/2}}\right\} = 0.95$$

$$\mathbb{P}\left\{\mu - 1.96\frac{\sigma}{n^{1/2}} \leq N(\mu, \sigma^2/n) \leq \mu + 1.96\frac{\sigma}{n^{1/2}}\right\} = 0.95$$

$$\mathbb{P}\left\{\mu - 1.96\frac{\sigma}{n^{1/2}} \leq \bar{X}_n \leq \mu + 1.96\frac{\sigma}{n^{1/2}}\right\} \approx 0.95$$

Building a Confidence Interval for the Expected Value of a Random Variable

- Therefore, we obtain the following interval that includes unknown μ with 0.95 probability

$$\mathbb{P}\left\{\bar{X}_n - 1.96\frac{\sigma}{n^{1/2}} \leq \mu \leq \bar{X}_n + 1.96\frac{\sigma}{n^{1/2}}\right\} = 0.95$$

- This computation requires knowing σ^2
- When σ^2 is not known, we estimate it by the sample standard deviation

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Building a Confidence Interval for the Expected Value of X

1. Select a sample size n
2. Generate n independent samples of X
3. Let these samples be $X_1,\ldots,X_n$
4. Compute

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

Building a Confidence Interval for the Expected Value of X

6. Look up the value of $z_{\alpha/2}$ such that

$$\mathbb{P}\left\{-z_{\alpha/2} \leq N(0,1) \leq z_{\alpha/2}\right\} = 1 - \alpha$$

7. E{X} lies in the interval

$$\bar{X}_n \mp z_{\alpha/2} \, \frac{s_n}{n^{1/2}}$$

approximately with probability 1-α

Gradient

- Let $f(x_1,\ldots,x_n)$ be a function from R^n to R
- The gradient of $f(x_1,\ldots,x_n)$ is the n-dimensional vector

$$\nabla f(x_1,\ldots,x_n) = \left[\frac{\partial f(x_1,\ldots,x_n)}{\partial x_1},\ldots,\frac{\partial f(x_1,\ldots,x_n)}{\partial x_n}\right]$$

- Gradient points towards the direction of steepest ascent

Gradient

- Gradient points towards the direction of steepest ascent
- Let the point $[y_1,\ldots,y_n]$ be obtained by

$$[y_1,\ldots,y_n] = [x_1,\ldots,x_n] + \alpha\,\nabla f(x_1,\ldots,x_n)$$

- If we choose α small enough, then

$$f(y_1,\ldots,y_n) \geq f(x_1,\ldots,x_n)$$

- Therefore, we can increase the value of the function by moving towards the gradient

Gradient

- Consider the function f(x,y) = 10 - x^2 - 4xy - $3y^2$

$$\nabla f(x,y) = [-2x - 4y, -4x - 6y]$$
$$f(1,1) = 2 \quad \nabla f(1,1) = [-6, -10]$$

- If we start from point [1,1] and move in direction [-6,-10] for a step size of 0.1, then we reach point

$$[1,1] + 0.1\,[-6,-10] = [0.4, 0] \quad f(0.4, 0) = 9.84$$

- If we start from point [1,1] and move in direction [-6,-10] for a step size of 0.5, then we reach point

$$[1,1] + 0.5\,[-6,-10] = [-2,-4] \quad f(-2,-4) = -74$$