EEL 4930 Stats – Lecture 24

CENTRAL LIMIT THEOREM

- Consider a sum of (independent) random variables:
- If $X_i,\ i=1,2,\ldots$ is a sequence of independent random variables with the same distribution and finite variance, then the distribution function of

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$$

  converges to a common distribution function

- This is the **Central Limit Theorem**– We won't cover the proof in this class – take EEE 5544
- The limiting distribution is that of a **Gaussian random variable**
- The density of a Gaussian RV X is

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

  which has two parameters: (mean) μ and (variance) $\sigma^2 \geq 0$

DISTRIBUTION FUNCTION

- The CDF of a Gaussian RV is given by

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt, \end{aligned}$$

  which cannot be evaluated in closed form

- Instead, we tabulate distribution functions for a normalized Gaussian variable with $\mu = 0$ and $\sigma^2 = 1$
- This is called the Normal distribution, and its CDF is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt$$

- Mathematicians use the "error function" (erf) to define the CDF of the normal distribution:

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right],$$

  where

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

- Engineers more commonly use the complementary distribution function, or Q-function, defined by

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\} dt$$

- Note that $Q(x) = 1 - \Phi(x)$

- I will be supplying you with a Q-function table and a list of approximations to $Q(x)$

- The Q-function can also be defined as

$$Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left\{-\frac{x^2}{\sin^2\phi}\right\} d\phi.$$

(This is a fairly recent result that is in very few textbooks. This form has a finite range of integration that is often easier to work with.)

- The CDF for a Gaussian RV with mean μ and variance σ^2 is

$$\begin{aligned} F_X(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right) \\ &= 1 - Q\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

Note that the denominator above is σ, not σ^2. Many students use the wrong value when solving problems!

- To find the prob. of some interval using the Q-function, it is easiest to rewrite the prob:

$$\begin{aligned} P(a < X \leq b) &= P(X > a) - P(X > b) \\ &= Q\left(\frac{a-\mu}{\sigma}\right) - Q\left(\frac{b-\mu}{\sigma}\right) \end{aligned}$$

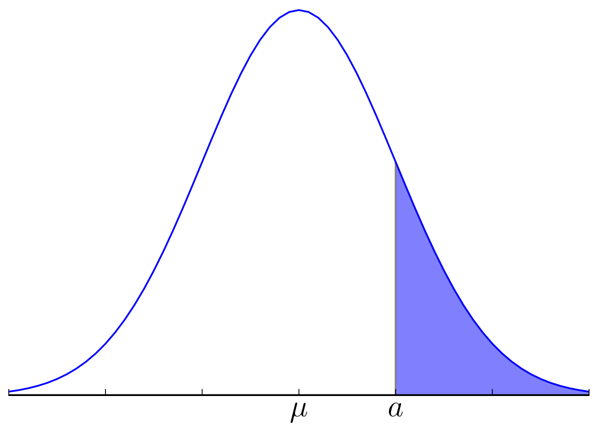
- **Engineering examples:** Noise sample in an electrical device, complex Gaussian models combined amplitude and phase of wireless signal received in multipath environment, sum of accumulated errors

1 More on Computing Gaussian Tail Probs

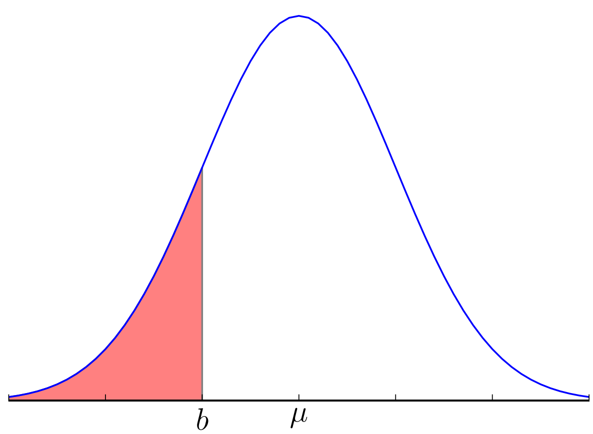
- Any Gaussian probabilities can be decomposed in terms of Gaussian tail probabilities

- There are 2 cases of the tail probabilities

- **Case 1:** $P(X \geq a)$, where $a > \mu$



- **Case 2:** $P(X \leq b)$, where $b < \mu$



EX Grading on a curve

A professor's class requests that he "grade on a curve". The professor sees that the class grades can be modeled using a Gaussian distribution with parameters μ and σ^2.

Let X represent a randomly chosen student's grade.

(a) What is the probability that the student's grade is above μ?

(b) The professor decides to use the following grading strategy:

- If the grades are within σ of the mean(μ), assign a B
- If the grades are more than σ above the mean, assign an A
- If the grades are more than σ below the mean, but less than 2σ below the mean, assign C
- If the grades are more than 2σ below the mean, but less than 3σ below the mean, assign D
- If the grades are more than 3σ below the mean, assign E

Determing the probability that a randomly chosen student gets each grade

(c) Suppose the threshold to get an A is $k\sigma$ above the mean, what value of k is needed for 40% of the class to get an A?