Lecture 5 (Ch. 1)

Last time we learned even more special distributions. So far, we have

(later) Related.

Bernoulli
Binomial
Poisson

$\left.\begin{matrix}\text{Bernoulli}\\ \text{Binomial}\\ \text{Poisson}\end{matrix}\right\} P(x) = \text{(probability) mass function (pmf)}$

discrete/categ.

$\left.\begin{matrix}\text{Uniform (in hw)}\\ \text{Exponential}\\ \text{Normal}\end{matrix}\right\} f(x) = \text{(probability) density function (pdf)}$

continuous.

The area "under" these dists is important, because it translates to the prob. that something can happen. e.g. $a<x<b$

Many of the areas are trivial to compute, but some are harder, and so they are tabulated. Check the Appendix in our book.

(Also note that props/probs are important and useful because they can be observed/measured in samples.)

Probs/areas for discrete/categ. variables are easier:

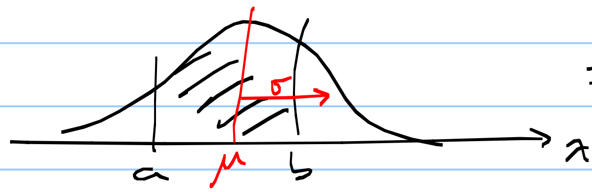
For example, if $x \sim \text{Binom}(n, \pi)$, then

$$\text{prob}(a \leq x \leq b) = \sum_{x=a}^{b} \underbrace{\boxed{\binom{n}{x} \pi^{x} (1-\pi)^{n-x}}}_{=P(x)}$$

Eventhough Table II gives some binomial areas, You do **NOT** have to use it. You can use this formula.

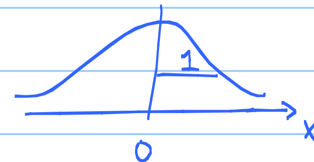
However, The normal dist. is one for whom areas are really hard to compute:

→ For Normal (μ, σ):



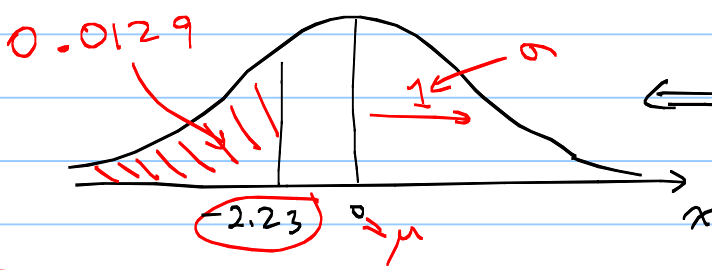
$$= \underbrace{\int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{= f(x) \text{ of Normal}} dx$$

→ For std. Normal: $\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx$



Note: std. Normal = Normal$(\mu=0, \sigma=1)$ = symmetric.

Unfortunately, integrals of This type can be done only numerically. Their values are tabulated in Table I. E.g.



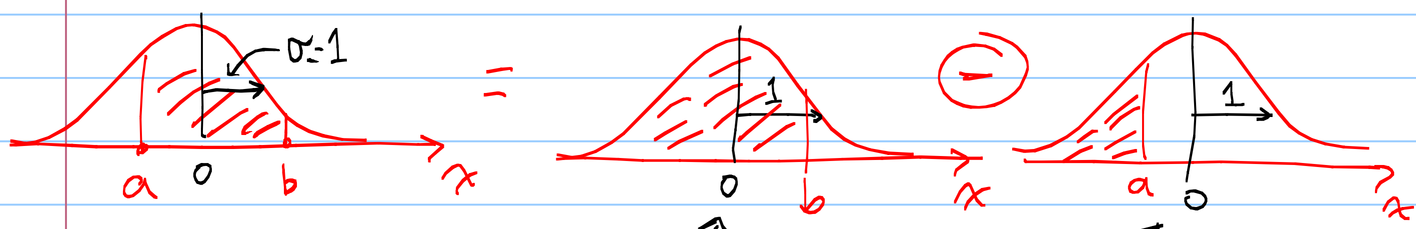
$$\Longleftrightarrow \int_{-\infty}^{-2.23} f(x)\, dx = 0.0129$$

$$\boxed{\text{In 390, use Table I, unless The problem says "By R."}}$$

So, now we know how to find area to the left of $x=a$, when x follows the std. Normal.

To find the area between 2 x's, there is a trick:

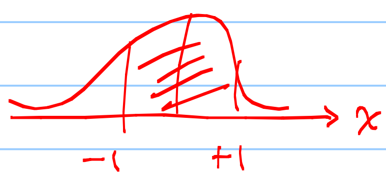
$\text{prob}(a < x < b) =$ area between a & b =



$= \text{prob}(x < b) - \text{prob}(x < a)$

Both of these can be obtained from Table I.

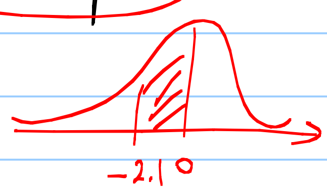
Example: What is the area under the std. Normal between −1 and +1?



$= .8413 - .1587 = 0.6826$

"famous" 68%

Example: How about between −2.1 and 0?



$= 0.5 - .0179 = \boxed{0.4821}$

Example: what is the area to the right of (−2.1)?

$$\left.\begin{array}{l} 1-(0.0179) \\ \text{or } 0.4821+0.5 \end{array}\right\} = 0.9821$$

Normal $(\mu=0, \sigma=1)$

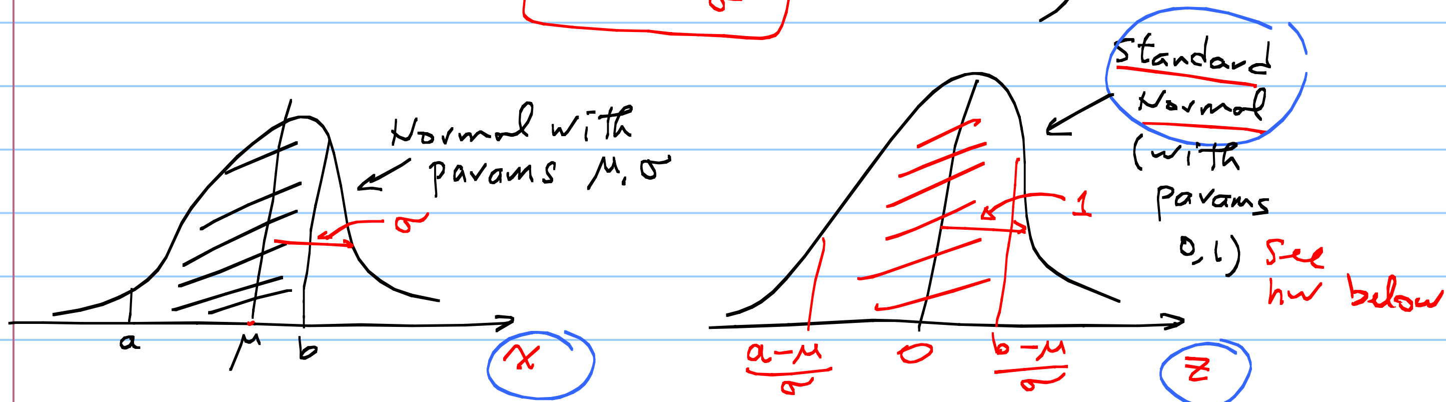
Now, we know how to find area/prop. under std. normal.

How do we handle $N(\mu, \sigma)$?

It would be impossible to tabulate values for every value of the 2 parameters, μ, σ. Need one more trick!

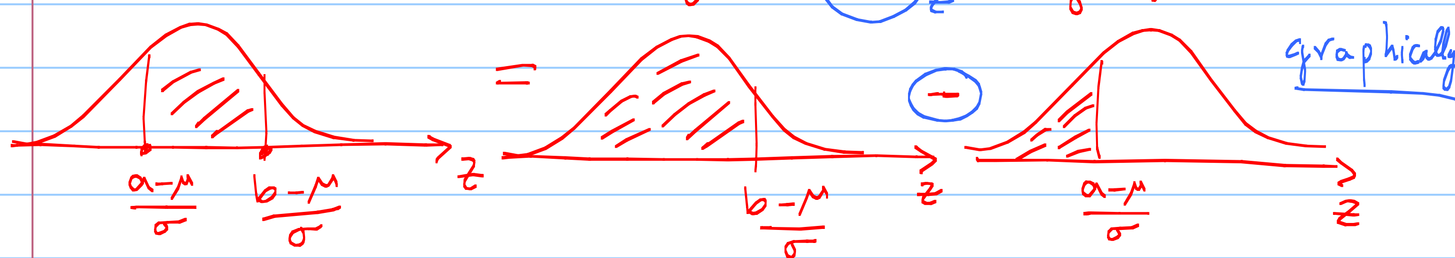
The trick is to "standardize" (ie. change variables):

$$x \longrightarrow \boxed{z = \frac{x-\mu}{\sigma}} \quad (z\text{-scores})$$



So, to compute area between 2 values:

$$\text{prob}(a < x < b) = \text{prob}\left(\frac{a-\mu}{\sigma} < \underbrace{\frac{x-\mu}{\sigma}}_{z} < \frac{b-\mu}{\sigma}\right)$$

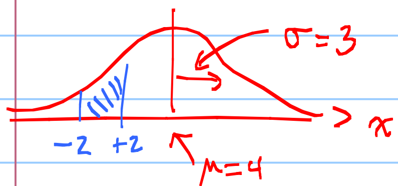


$$= \text{prob}\left(z < \frac{b-\mu}{\sigma}\right) - \text{prob}\left(z < \frac{a-\mu}{\sigma}\right)$$

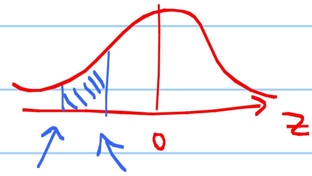
Algebraically

Either way (algebraically or graphically) you can obtain the value of each term from Table 1.

Example: What's the area between -2 and $+2$ for a normal curve with $\mu = 4$, $\sigma = 3$



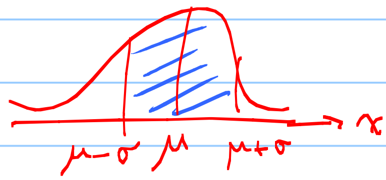
=



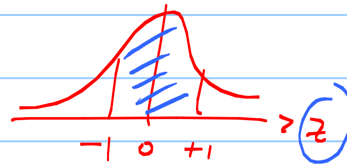
$= 0.2514 - 0.0228 = \boxed{0.2286}$

$-2 \Rightarrow \frac{-2-4}{3}$, $\quad \frac{2-4}{3} = -0.67$

Example: What's the prob. of x being within 1σ of μ?



=



$= .8413 - .1587 = \boxed{0.6826}$

Example: The prob. of being beyond of 1σ of μ?

$1 - 0.6826 = \boxed{0.3174}$

Summary:

Given $f(x)$, and $x = a$ (and/or b), we can compute area.

If $f(x) =$ std. Normal, Then Table I.

$\hookrightarrow x \sim N(0,1)$

If $f(x) = \text{Normal}(\mu, \sigma)$, Then $\boxed{\text{standardize first}}$, and proceed ...

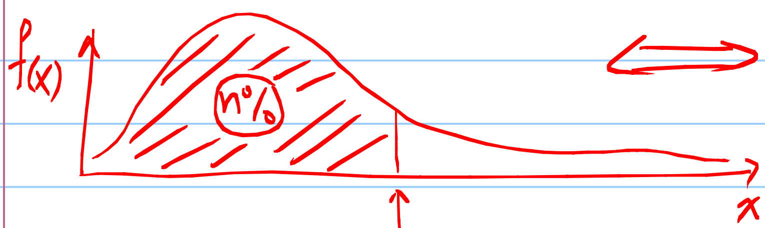
$\hookrightarrow x \sim N(\mu, \sigma)$ $\qquad z = \frac{x-\mu}{\sigma}$

Let's turn things around: Given f(x), and area, find x.

E.g. median: $\int_{-\infty}^{\text{median}} f(x)\,dx = 1/2$

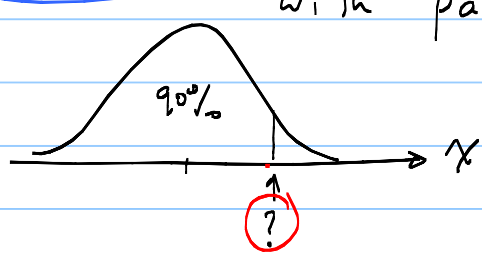
median = 50^{th} percentile = 0.5 quantile = 2^{nd} quartile

So, median is a special case of a more general concept:



n^{th} percentile $\Longleftrightarrow$ $\int_{-\infty}^{\boxed{n^{th}\text{ percentile}}} f(x)\,dx = \frac{n}{100}$ $\boxed{\text{Defn}}$

Example: What's the 90^{th} percentile of a normal distr. with params μ, σ?



$\Longrightarrow$

$\frac{?-\mu}{\sigma} = 1.285$ ← Table I

$? = (\mu + 1.285\sigma)$

> Note: percentile is a number on the x-axis, not a percent.
> I.e. a percentile of x has the same units as x.

By now, you should be able to (for hists AND dists)

1) compute the area to the left (or right) of $x=a$,
2) compute " " between $x=a$, $x=b$,
3) compute $x=a$, given the area to left (or right),

If the left area is $n\%$, then $x=a$ is called the n^{th} percentile.

Q1: Which of The following can NOT be a percentile of some variable?

a) 183   b) -1.3 c) 10% $\boxed{\text{d) None of The above.}}$

a) e.g. height in cm.

b) e.g. temperature in Centigrade

c) e.g. grades in percent.

d) (ie. all of These can be a percentile)

---

hw- lect 5-1

Suppose The density function for x is given by The Normal dist. with parameters μ, σ. I.e.

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

a) Compute The density function, $f(z)$, for $z = \frac{x-\mu}{\sigma}$.

<u>Important Hint:</u> $f(z)$ must satisfy $\int_{-\infty}^{\infty} f(z)\,dz = 1$.

So, start with $\int_{-\infty}^{\infty} f(x)\,dx = 1$, with $f(x)$ = above, and massage The expression until it becomes $\int_{-\infty}^{\infty} [\ldots]\,dz = 1$. Then $f(z) = [\ldots]$.

Note: It is <u>not</u> necessary to perform any integrals.

b) In The place where μ and σ appear in $f(x)$, what values do you find in $f(z)$?

hw- lect 5-2

What's The 10^{th} percentile of The uniform dist. between $-1, +1$?

Hint: for uniform dist. integration is Trivial.

hw- lect 5-3

Find The n^{th} percentile of an exponential dist. with param. λ.

Hint: The answer will depend on λ and n.