Christopher Lum
lum@uw.edu

Lecture 09c
Normal/Gaussian Distributions

Lecture is on YouTube

The YouTube video entitled ‘Gaussian/Normal Distributions’ that covers this lecture is located at https://youtu.be/Xaju4l9KTE0.

Outline

-Normal Distribution

Normal Distribution

Turning from discrete to continuous distributions, we now consider the most important continuous distributions, the **normal distribution** (aka **Gaussian distribution**). The probability density of this distribution is defined as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \qquad \textbf{\textit{(Eq.1)}}$$

where $\sigma > 0$ = standard deviation

Some features of $f(x)$ include

1. μ is the mean and σ is the standard deviation
2. $\frac{1}{\sigma\sqrt{2\pi}}$ is a constant factor that makes the area under the curve from $-\infty < f(x) < \infty$ equal to 1 as required by Eq.10 in section 24.5. This is also sometimes referred to as the normalizer.
3. The curve of $f(x)$ is symmetric with respect to $x = \mu$.
4. The exponential function in Eq.1 goes to zero very fast. Smaller standard deviations lead to faster decay away from $x = \mu$.

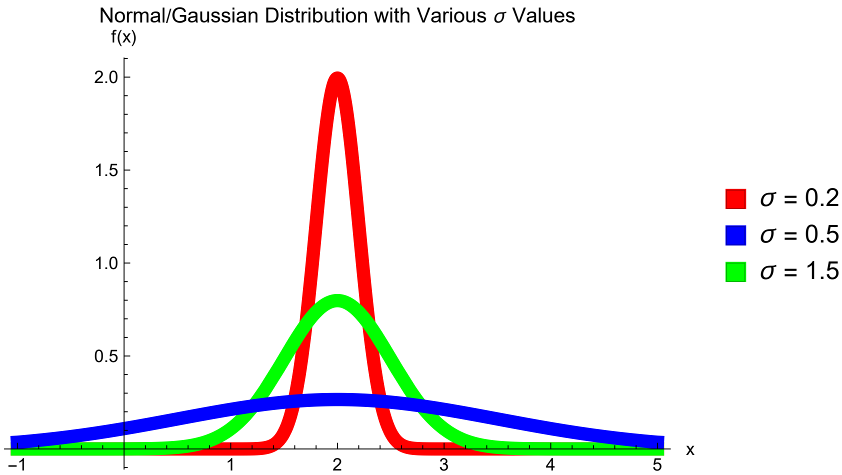
Note that in some situations, we refer to the variance instead of the standard deviation. The variance is simply the standard deviation squared.

variance = σ^2

We can plot several situations

```
(*Define the Normal distribution*)
f[x_, μ_, σ_] = 1/(σ Sqrt[2 π]) Exp[-1/2 ((x - μ)/σ)^2]

(*Plot options*)
deltaX = 3;
μGiven = 2;

σ1 = 0.2;
σ2 = 0.5;
σ3 = 1.5;
Legended[
 Show[
  (*Plot 1*)
  Plot[f[x, μGiven, σ1], {x, μGiven - deltaX, μGiven + deltaX},
   PlotStyle → {Red, Thickness[0.02]}, PlotRange → All],

  (*Plot 2*)
  Plot[f[x, μGiven, σ2], {x, μGiven - deltaX, μGiven + deltaX},
   PlotStyle → {Green, Thickness[0.02]}, PlotRange → All],

  (*Plot 3*)
  Plot[f[x, μGiven, σ3], {x, μGiven - deltaX, μGiven + deltaX},
   PlotStyle → {Blue, Thickness[0.02]}, PlotRange → All],

  (*Plot Options*)
  PlotLabel → "Normal/Gaussian Distribution with Various σ Values",
  AxesLabel → {"x", "f(x)"}
 ],

 (*Add legend information*)
 SwatchLegend[
  {Red, Blue, Green}, {
   StringJoin["σ = ", ToString[σ1]],
   StringJoin["σ = ", ToString[σ2]],
   StringJoin["σ = ", ToString[σ3]]
  }
 ]
]

Clear[σ3, σ2, σ1, μGiven, deltaX]
```

$$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$$



This is sometimes referred to as a “Bell Curve” due to the bell like shape of $f(x)$ vs. x. However it should be noted that other distributions such as the Cauchy, Student-T, and Logistic distribution also display similar “bell-like” shapes.

Mathematica provides the ‘NormalDistribution’ object to represent this.

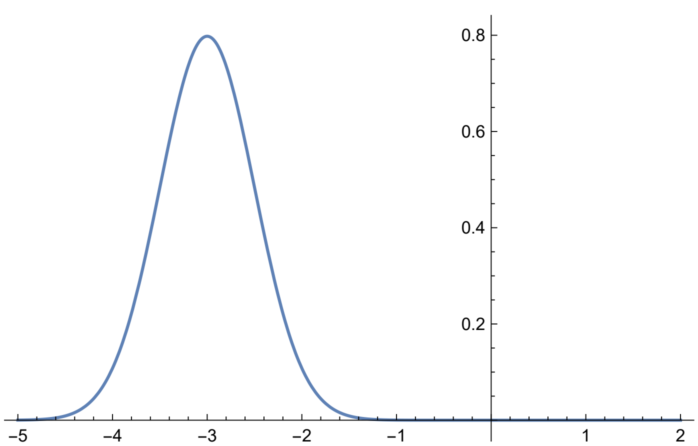
```
NormalDistribution[μ, σ]
```

```
NormalDistribution[μ, σ]
```

We see that by itself, the ‘NormalDistribution’ object is not terribly useful. However we can use it in conjunction with the ‘PDF’ function to obtain an expression for the distribution

```
f2[x_, μ_, σ_] = PDF[NormalDistribution[μ, σ], x]
Plot[f2[x, -3, 1 / 2], {x, -5, 2}]
Clear[f2]
```

$$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$$



Erf (Error Function)

Consider a Gaussian distribution with mean of zero ($\mu = 0$) and variance of 1/2 (and therefore a standard deviation of $\sigma = \sqrt{1/2}$).

```
f[x, 0, (1 / 2)^(1/2)]
```

$$\frac{e^{-x^2}}{\sqrt{\pi}}$$

So we have

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

If we would like to compute the probability that a sample drawn from this distribution falls between $[-x, x]$, we can integrate this between $-x$ and x

$$p(z \in [-x, x]) = \int_{-x}^{x} f(t)\, dt$$

We note that since the integrand is an even function ($f(-x) = f(x)$), integrating from $-x$ to x is the same as integrating from 0 to x and then multiplying the result by 2 (show graphical representation) so we can write

$$= 2 \int_0^x f(t)\, dt$$

$$p(z \in [-x, x]) = 2 \int_0^x \left(\frac{1}{\sqrt{\pi}} \exp\left(-t^2\right)\right) dt$$

This quantity (the probability that a sample drawn from a Gaussian with $\mu = 0$ and $\sigma = \sqrt{1/2}$ falls within -x and x) is known as the error function, erf

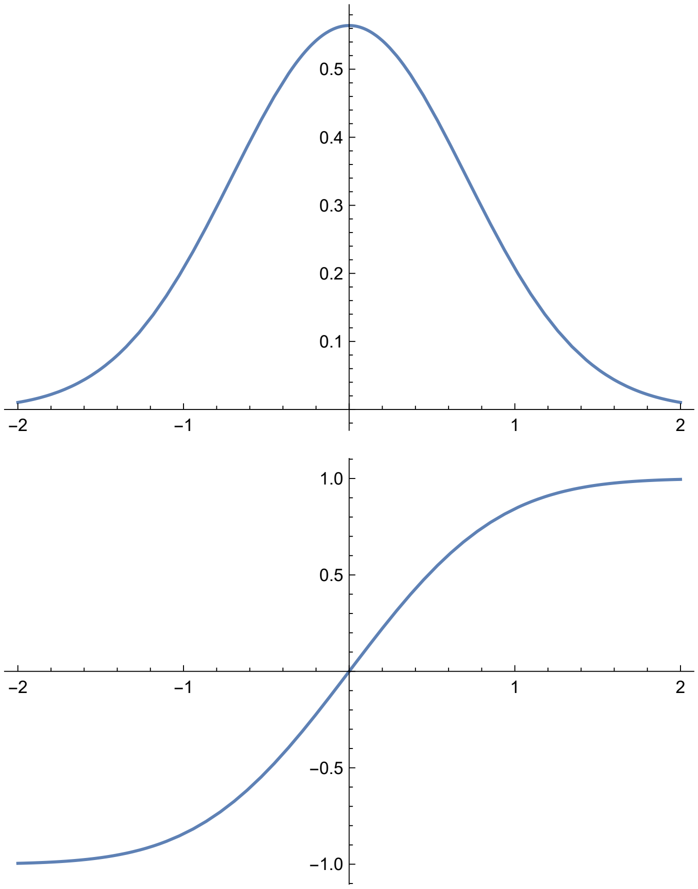
$$\mathrm{erf}(x) = 2 \int_0^x \left(\frac{1}{\sqrt{\pi}} \exp\left(-t^2\right)\right) dt$$

```
Integrate[f[t, 0, (1 / 2)^(1/2)], {t, -x, x}]
```

```
Erf[x]
```

Because this integral cannot be computed analytically, many computer languages provide numerical tables for the error function. Mathematica has the ‘Erf’ function to compute this function. Matlab provides the ‘erf’ function to do the same operation.

```
Plot[f[x, 0, (1 / 2)^(1/2)], {x, -2, 2}]
Plot[Erf[x], {x, -2, 2}]
```



We can now use the error function to describe cumulative distribution functions of arbitrary Gaussian distributions.

Cumulative Distribution Function $F(x)$

Recall from Eq.7 in section 24.5 that the cumulative distribution function, $F(x)$, is related to the probability density function, $f(x)$, through

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$ ***(Eq.7 from section 24.5)***

So for the case of a generalized normal distribution, we have

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt$$ ***(Eq.2)***

We cannot easily obtain an analytical expression for this integral.

```
F[x_, μ_, σ_] = Simplify[
  Integrate[f[t, μ, σ], {t, -∞, x}],
  {σ > 0}]
```

$$\frac{1}{2} \times \left(1 + \mathrm{Erf}\left[\frac{x - \mu}{\sqrt{2}\ \sigma}\right]\right)$$

We can also attempt to use Mathematica's 'CDF' function in conjunction with the 'NormalDistribution' function to evaluate Eq.2

```
F2[x_, μ_, σ_] = CDF[NormalDistribution[μ, σ], x]
```

$$\frac{1}{2}\ \mathrm{Erfc}\left[\frac{-x + \mu}{\sqrt{2}\ \sigma}\right]$$

In this case, note that Erfc is the complementary error function which is related to the normal error function through

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$$

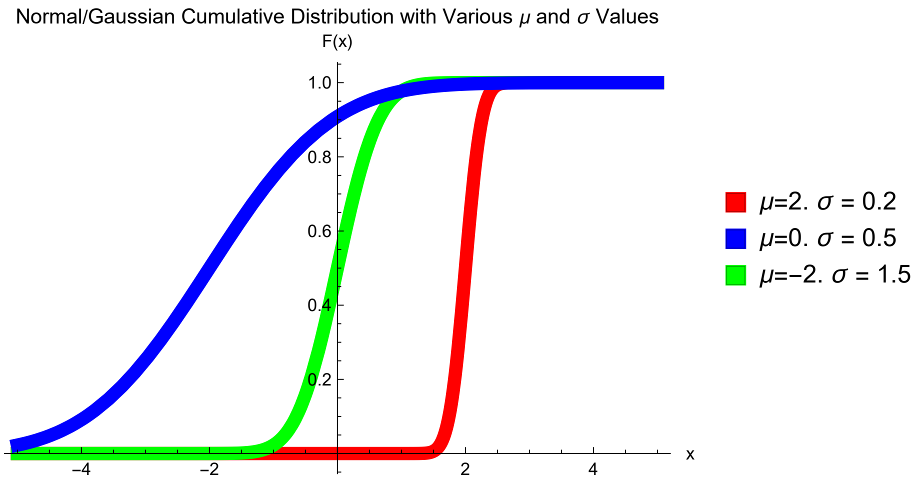
With this knowledge, the previous two expressions for $F(x)$ can be shown to be the same expression

```
FullSimplify[F[x, μ, σ] == F2[x, μ, σ]]
```

```
True
```

We can now plot $F(x)$ for various μ and σ

```
(*Plot options*)
xMin = -5;
xMax = 5;

μ1 = 2; σ1 = 0.2;
μ2 = 0; σ2 = 0.5;
μ3 = -2; σ3 = 1.5;

Legended[
 Show[
  (*Plot 1*)
  Plot[F[x, μ1, σ1], {x, xMin, xMax},
   PlotStyle → {Red, Thickness[0.02]}, PlotRange → All],

  (*Plot 2*)
  Plot[F[x, μ2, σ2], {x, xMin, xMax},
   PlotStyle → {Green, Thickness[0.02]}, PlotRange → All],

  (*Plot 3*)
  Plot[F[x, μ3, σ3], {x, xMin, xMax},
   PlotStyle → {Blue, Thickness[0.02]}, PlotRange → All],

  (*Plot Options*)
  PlotLabel → "Normal/Gaussian Cumulative Distribution with Various μ and σ Values",
  AxesLabel → {"x", "F(x)"}
 ],

 (*Add legend information*)
 SwatchLegend[
  {Red, Blue, Green}, {
   StringJoin["μ=", ToString[μ1], ". σ = ", ToString[σ1]],
   StringJoin["μ=", ToString[μ2], ". σ = ", ToString[σ2]],
   StringJoin["μ=", ToString[μ3], ". σ = ", ToString[σ3]]
  }
 ]
]

Clear[σ3, σ2, σ1, μ1, μ2, μ3, deltaX, xMin, xMax]
```



If $\mu = 0$ and $\sigma = 1$, we have the **standard normal distribution** which is typically denoted by ϕ (lower case phi)

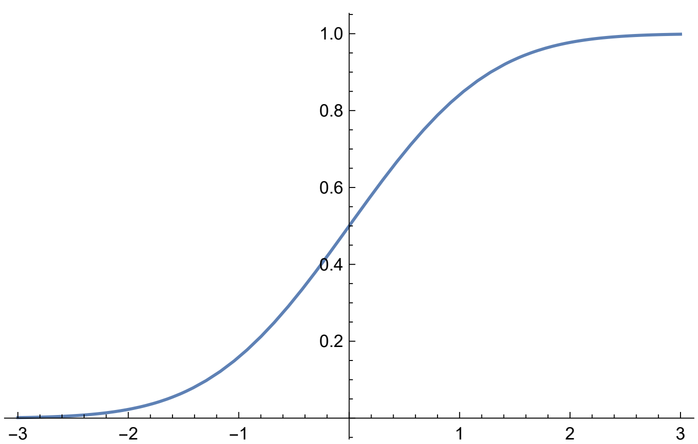
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} x^2\right]$$

The cumulative distribution function of the standard normal distribution yields the **standard cumulative distribution function** and is typically denoted by $\Phi(x)$ (capital phi)

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

```
Φ[x_] = F[x, 0, 1]
Plot[Φ[x], {x, -3, 3}]
```

$$\frac{1}{2} \times \left(1 + \text{Erf}\left[\frac{x}{\sqrt{2}}\right]\right)$$



So we have

$$\Phi(x) = \frac{1}{2} \times \left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \qquad \textbf{\textit{(Eq.3)}}$$

The benefit of Eq.3 is that if we have an implementation of the error function (as provided by many different computer programs), we can compute the standard cumulative distribution function Φ

The standardized cumulative distribution function is related to the general cumulative distribution distribution with any μ and σ through the following theorem.

Theorem 1: The cumulative distribution function $F(x)$ of the normal distribution with any μ and σ is related to the standardized cumulative distribution function Φ by

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \textbf{(Eq.4)}$$

where $F(x)$ = cumulative distribution function of a normal/Gaussian with mean μ and std dev σ

We can verify this through direct substitution or alternatively see the textbook for proof.

```
F[x, μ, σ] == Φ[(x - μ)/σ] // Simplify
```

```
True
```

With this information in hand, we obtain the probability that a sample from an arbitrary Gaussian will assume any value within a given interval

Theorem 2: Normal Probabilities for Intervals

The probability that a normal random variable X with mean μ and standard deviation σ will assume any value in an interval $a < x \leq b$ is

$$P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad \textbf{(Eq.5)}$$

The proof of Eq.5 is simply applying the direct substitution we applied earlier to validate Eq.4

Eq.5 gives us the ability to compute $P(a < X \leq b)$ using the error function. In other words, Theorem 1 can alternatively be stated as

Theorem 1 (Using Erf): The distribution function $F(x)$ of the normal distribution with any μ and σ is related to the error function by

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2} \times \left(1 + \operatorname{erf}\left(\frac{1}{\sqrt{2}} \frac{x-\mu}{\sigma}\right)\right) \qquad \textbf{(Eq.A.3)}$$

```
Φ[(x - μ)/σ] == 1/2 × (1 + Erf[1/√2 (x - μ)/σ]) // FullSimplify
F[x, μ, σ] == 1/2 × (1 + Erf[1/√2 (x - μ)/σ]) // FullSimplify
```

```
True
```

```
True
```

In a similar fashion, the probability that X is in an interval was already given in Theorem 2 which can now be restated using the error function

Theorem 2 (Using Erf): Normal Probabilities for Intervals

The probability that a normal random variable X with mean μ and standard deviation σ will assume any value in an interval $a < x \leq b$ is

$$P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{b-\mu}{\sigma}\right)\right) - \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{a-\mu}{\sigma}\right)\right) \qquad \textbf{\textit{(EqA.4)}}$$

Example 2 in Textbook But Use Erf Instead of Φ

Let X be normal with mean 0.8 and variance 4 (so that $\sigma = 2$). We can compute various probabilities

i. $P(X \leq 2.44)$
ii. $P(X \geq 1)$
iii. $P(1.0 \leq X \leq 1.8)$

<u>Use standard cumulative normal distribution function, Φ</u>

i. $P(X \leq 2.44) = F(2.44) = \Phi\left(\frac{2.44-0.8}{2}\right) = \Phi(0.82) = 0.7939$
ii. $P(X \geq 1) = 1 - P(X \leq 1) = 1 - \Phi\left(\frac{1-0.8}{2}\right) = 1 - 0.5398 = 0.4602$
iii. $P(1.0 \leq X \leq 1.8) = \Phi\left(\frac{1.8-0.8}{2}\right) - \Phi\left(\frac{1.0-0.8}{2}\right) = \Phi(0.5) - \Phi(0.1) = 0.1517$

<u>Which is practically implemented using the error function, erf</u>

i. $P(X \leq 2.44) = F(2.44) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{2.44-0.8}{2}\right)\right) = 0.7939$
ii. $P(X \geq 1) = 1 - P(X \leq 1) = 1 - \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{1-0.8}{2}\right)\right) = 0.4602$
iii. $P(1.0 \leq X \leq 1.8) = \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{1.8-0.8}{2}\right)\right) - \frac{1}{2} \times \left(1 + \mathrm{erf}\left(\frac{1}{\sqrt{2}} \frac{1.0-0.8}{2}\right)\right) = 0.1516$

```
Print["part i"]
```
$$\frac{1}{2}\times\left(1+\text{Erf}\left[\frac{1}{\sqrt{2}}\left(\frac{2.44-0.8}{2}\right)\right]\right)$$
```
F[2.44, 0.8, 2]
Print[]

Print["part ii"]
```
$$1-\frac{1}{2}\times\left(1+\text{Erf}\left[\frac{1}{\sqrt{2}}\left(\frac{1-0.8}{2}\right)\right]\right)$$
```
1 - F[1, 0.8, 2]
Print[]

Print["part iii"]
```
$$\frac{1}{2}\times\left(1+\text{Erf}\left[\frac{1}{\sqrt{2}}\left(\frac{1.8-0.8}{2}\right)\right]\right)-\frac{1}{2}\times\left(1+\text{Erf}\left[\frac{1}{\sqrt{2}}\left(\frac{1-0.8}{2}\right)\right]\right)$$
```
F[1.8, 0.8, 2] - F[1, 0.8, 2]
Print[]
```

```
part i
0.793892
0.793892


part ii
0.460172
0.460172


part iii
0.151635
0.151635
```

As expected, this also works with distributions with negative mean. For example $\mu = -120.5$, $\sigma = 50$

```
(*what is probability that sample is less than the mean (and mean is negative)*)
μNegative = -120.5;
F[μNegative, μNegative, 50]
```

```
0.5
```

Numeric Values

In practical work with the normal distribution, it is good to remember that about 2/3 of all values of X to be observed with lie between $\mu \pm \sigma$, about 95% between $\mu \pm 2\,\sigma$, and practically all between the

three-sigma limit $\mu \pm 3\,\sigma$. We can compute the precise probabilities using Eq.5.

```
Print["μ±σ"]
F[μ + σ, μ, σ] - F[μ - σ, μ, σ] // N
Print[]

Print["μ±2σ"]
F[μ + 2 σ, μ, σ] - F[μ - 2 σ, μ, σ] // N
Print[]

Print["μ±3σ"]
F[μ + 3 σ, μ, σ] - F[μ - 3 σ, μ, σ] // N
Print[]
```

μ±σ

```
0.682689
```

μ±2σ

```
0.9545
```

μ±3σ

```
0.9973
```

Graphically, we recognize this as the area under the probability density function, $f(x)$ as shown in Figure 521.

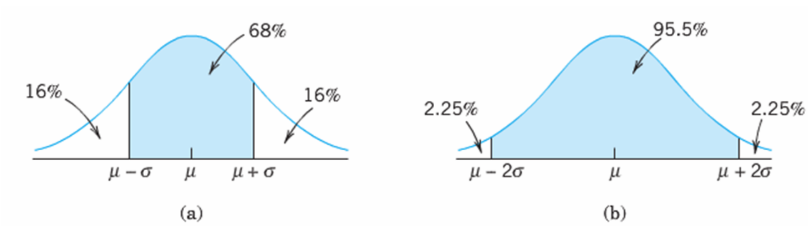
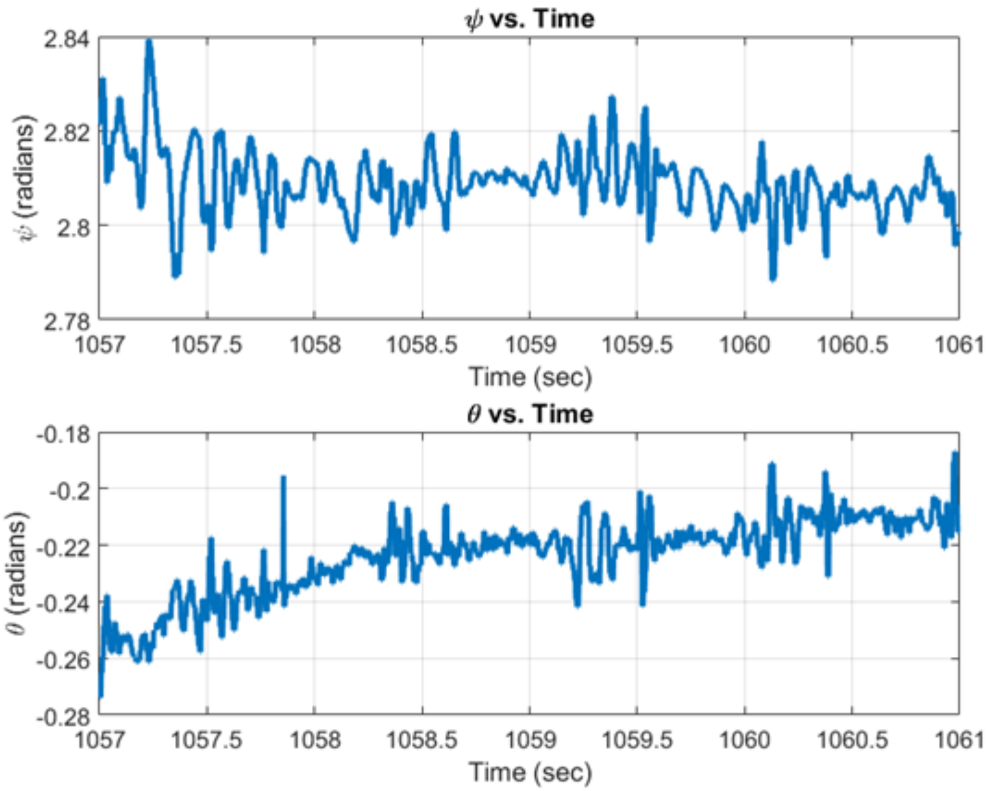


Fig. 521. Illustration of formula (6)

Real Data Modeled as a Gaussian

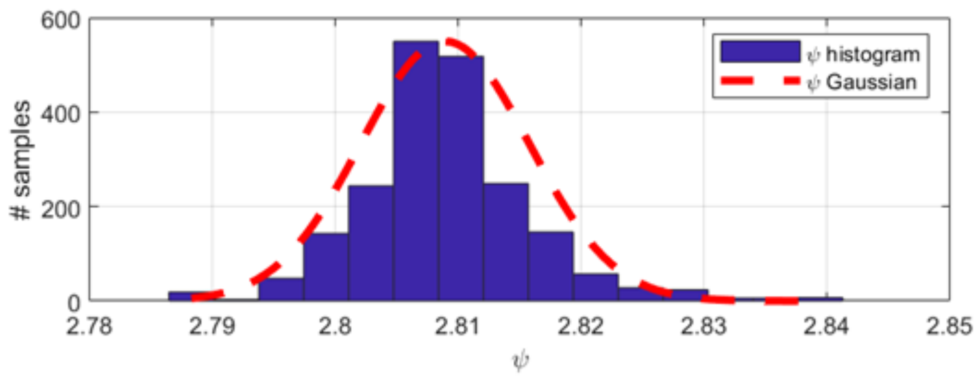
Consider the data set below

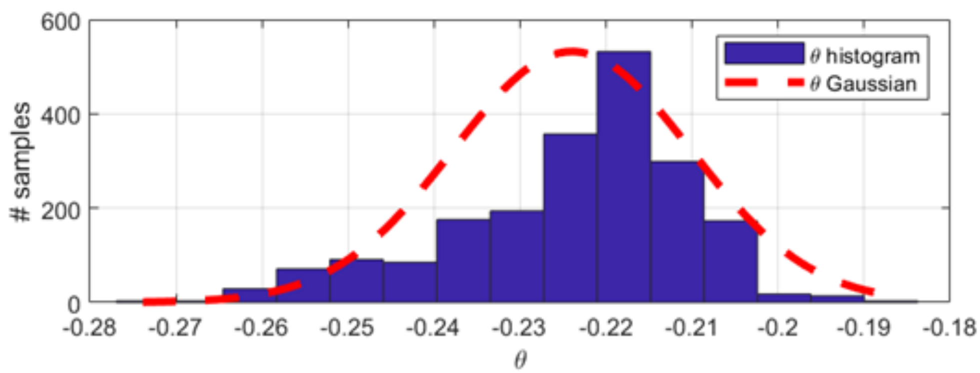


We can compute a Gaussian distributions to fit these samples, for more information see the Matlab code (result shown below). Of particular note is that

$\mu = 2.8089$ (mean)

$\sigma = 0.0069$ (standard deviation)





We note that this fits quite nicely for the ψ distribution but since the θ measurement appears to have a changing mean, the fitted gaussian does not match the samples quite as well but is still a reasonable approximation.

Sampling From a Normal Distribution

It may be necessary to sample from a normal distribution.

Talk about 'rand' vs. 'randn' . Note that 'randn' samples from a standard Normal distribution ($\mu = 0$ and $\sigma = \sigma^2 = 1$). We can transform to an arbitrary Gaussian by applying the results from the previous lecture which is repeated here for convenience.

Theorem 2: Transformation of Mean and Variance

If a random variable X has mean μ and variance σ^2, then the random variable

$$X^* = a_1 + a_2 X \qquad \textbf{(Eq.4)}$$

has mean μ^* and σ^{*2} where

$$\mu^* = a_1 + a_2 \mu \qquad \textbf{(Eq.5)}$$
$$\sigma^{*2} = a_2{}^2 \sigma^2$$

In this situation, it would be useful to solve for a_1 and a_2

```
Simplify[Solve[{μStar == a1 + a2 μ, varStar == a2^2 var}, {a1, a2}], {var > 0, varStar > 0}]
```

$$\left\{\left\{a1 \to \sqrt{\frac{varStar}{var}}\ \mu + \mu Star,\ a2 \to -\sqrt{\frac{varStar}{var}}\right\},\ \left\{a1 \to -\sqrt{\frac{varStar}{var}}\ \mu + \mu Star,\ a2 \to \sqrt{\frac{varStar}{var}}\right\}\right\}$$

Choosing the second solution yields

$$a_1 = -\sqrt{\frac{\sigma^{*2}}{\sigma^2}}\ \mu + \mu^* \qquad a_2 = \sqrt{\frac{\sigma^{*2}}{\sigma^2}}$$

or

$$a_1 = -\frac{\sigma^*}{\sigma}\ \mu + \mu^* \qquad a_2 = \frac{\sigma^*}{\sigma}$$

In the case of X being sampled from a standard gaussian, we have $\mu = 0$ and $\sigma = 1$ so this simplifies to

$$a_1 = \mu^* \qquad a_2 = \sigma^* \qquad (\text{if } \mu = 0 \text{ and } \sigma = 1 \text{ only})$$

For example, if we want to transform our set of samples from a standard normal distribution to a set of samples from a distribution with $\mu^* = -15.5$ and $\sigma^* = 3.5$, we use

$$a_1 = -15.5 \qquad a_2 = 3.5$$

Central Limit Theorem

The Central Limit Theorem states that in some situations, the distribution of sample means will approximate a normal distribution regardless of the distribution used to generate the samples.

Consider a sample of size N (in other words the sample contains N observations)

$$X = \{X_1, X_2, \ldots, X_N\}$$

This sample (and therefore the N observations) is drawn from a distribution with mean μ and variance σ^2.

Suppose we calculate the sample average of X and denote this as S_N

$$S_N = \frac{X_1 + X_2 + \ldots + X_n}{N}$$

It should be no surprise that as $N \to \infty$, $S_N \to \mu$. What is surprising though is if we repeat this for M samples, that the distribution of the M values of S_N is a normal distribution with mean $\mu_X = \mu$ and standard deviation $\sigma_X = \sigma / \sqrt{N}$.

Example: Sum of Uniform Distributions

Consider a uniform distribution between $a = 2$ and $b = 5$.

From our previous lecture, we know that the mean and variance of this distribution is given by

$$\mu = \frac{a+b}{2} = 3.5$$

$$\sigma^2 = \frac{1}{12}\,(a-b)^2 = 0.75$$

Study A: $n = 4$, $M = 3$

The 3 instances of samples are shown below

```
X =

    4.4442    4.7174    2.3810    4.7401


X =

    3.8971    2.2926    2.8355    3.6406


X =

    4.8725    4.8947    2.4728    4.9118
```

The 3 values of S_N are therefore

$S_4 = 4.0707$ (sample 1)
$S_4 = 3.1665$ (sample 2)
$S_4 = 4.2880$ (sample 3)

We can then plot this

Normal Approximation of the Binomial Distribution

Recall that the probability function of the binomial distribution is

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad (x = 0, 1, \ldots, n) \qquad \textbf{\textit{(Eq.8)}}$$

If n is large, the binomial coefficients and power become very inconvenient. It is of great practical (and theoretical) importance that, in this case, the normal distribution provides a good approximation of the binomial distribution, according to the following theorem, one of the most important theorems in all probability theory.

Theorem 3: Limit Theorem of De Moivre and Laplace
For large n

$$f(x) \sim f^*(x) \quad (x = 0, 1, \ldots, n) \qquad \textbf{\textit{(Eq.9)}}$$

where $\quad f(x) = \binom{n}{x} p^x q^{n-x} \quad$ (binomial distribution)

$$f^*(x) = \frac{1}{\sqrt{2\pi}\,\sqrt{n\,p\,q}}\, e^{-z^2/2}, \quad z = \frac{x - n\,p}{\sqrt{n\,p\,q}} \qquad \textbf{\textit{(Eq.10)}}$$

Here, $f^*(z)$ is the density of the normal distribution with mean $\mu = n\,p$ and variance $\sigma^2 = n\,p\,q$ (the mean and variance of the binomial distribution).

The symbol ~ (read asymptotically equal) means that the ratio of both sides approaches 1 as n approaches ∞.

Furthermore, for any non-negative integers a and $b > a$

$$P(a \le X \le b) = \sum_{x=a}^{b} \binom{n}{x} p^x q^{n-x} \sim \Phi(\beta) - \Phi(\alpha) \qquad \textbf{\textit{(Eq.11)}}$$

where $\quad \alpha = \frac{a - n\,p - 1/2}{\sqrt{n\,p\,q}}$

$$\beta = \frac{b - n\,p + 1/2}{\sqrt{n\,p\,q}}$$