Hawes 1

# Lecture #25 Normal Distribution, Transformations, and Statistics

## I. Gauss' Normal Distribution

### A. Basics

1. Definition Gauss Distribution

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}} \qquad -\infty < x < \infty$$

a. Mean: $\mu$ Variance: $\sigma^2$

2. Conditional Probability

a. $P(|x-\langle x\rangle| > k\sigma) = P(|y| > k)$ $\qquad y = \frac{x-\langle x\rangle}{\sigma}$

$$= \frac{\sqrt{2}}{\sqrt{\pi}}\int_k^\infty e^{-\frac{y^2}{2}} dy = \sqrt{\frac{4}{\pi}}\int_{k/\sqrt{2}}^\infty e^{-z^2} dz \equiv \mathrm{erfc}\left(\frac{k}{\sqrt{2}}\right)$$

Complimentary error func.

$z = \frac{y}{2}$

b. Note that $P(|x-\langle x\rangle| \geq 3\sigma) = 0.0027$

Much, much less than Chebyshev inequality $(\frac{1}{9})$ for abitrary distribution

3. Relation to Poisson Distribution, $p(n) = \frac{\mu^n}{n!} e^{-\mu}$

a. Theorem: For large $n$ and mean value $\mu$, the poisson distribution approaches a Gauss distribution.

b. Can be proven using deviation $V \equiv n - \mu$ as new variable, taking $\mu \to \infty$, $V/\mu$ small, and $V^2/\mu$ finite. Makes use of Stirling's Formula for $n!$ and $n \gg 1$ : $\boxed{n! \simeq \sqrt{2n\pi}\left(\frac{n}{e}\right)^n}$

c. Shows that $p(n) \sim \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{V^2}{2\mu}}$, Gauss Dist with $\sigma = \sqrt{\mu}$

I. A. (Continued)

4. Relation to Binomial Distribution:

a. Theorem: In the limit $n \to \infty$ with $p$ a finite probability such that $pn \to \infty$, the binomial distribution becomes a Gauss distribution.

## II. Transformations of Random Variables

### A. General Transformations

1. We have already seen, for $Y = aX + b$, $\langle Y \rangle = a\langle X \rangle + b$, $\sigma^2(Y) = a^2\sigma^2(X)$

2. Transformation: $y = y(x)$ from $X$ to $Y$:

a. For $X$ with probability $f(x)$, $Y$ with probability $g(y)$,

$$g(y)dy = f(x)dx \quad \text{or} \quad \boxed{g(y) = f[x(y)]\frac{dx}{dy}}$$

b. Ex: i) $y = x^2 \Rightarrow x = \sqrt{y}$

ii) $\frac{dx}{dy} = \frac{1}{2}y^{-\frac{1}{2}}$

iii) thus $\boxed{g(y) = f(\sqrt{y})\frac{1}{2\sqrt{y}}}$

3. Transformation from $X, Y$ to $U(x,y)$, $V(x,y)$:

a. $U = U(x,y)$ $V = V(x,y)$ $\Rightarrow x = x(U,V)$, $y = y(U,V)$

b. $$\boxed{g(U,V) = f[x(U,V), y(U,V)]\,|J|}$$ Jacobian

c. Jacobian: $$J = \frac{\partial(x,y)}{\partial(U,V)} = \begin{vmatrix} \frac{\partial x}{\partial U} & \frac{\partial x}{\partial V} \\ \frac{\partial y}{\partial U} & \frac{\partial y}{\partial V} \end{vmatrix}$$

II. (Continued)

## B. Addition of Random Variables

1. Consider $Z = X + Y$

   a. Transform from $X, Y$ to $X, Z$

   b. $x = x$, $z = x + y$, $\iff$ $x = x$ $\quad y = z - x$

   c. $$J = \begin{vmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial z} \\ \frac{\partial (z-x)}{\partial x} & \frac{\partial (z-x)}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

   d. Thus $\boxed{g(x,z) = f(x, z-x)}$

2. If we want marginal distribution, $h(z) = \int g(x,z)\,dx$

   a. If $X$ & $Y$ are independent, then $f(x,y) = f_1(x) f_2(y)$

   b. Thus $$\boxed{h(z) = \int_{-\infty}^{\infty} f_1(x) f_2(z-x)\,dx}$$

   c. NOTE: This is the Fourier convolution, $h(z) = \sqrt{2\pi}\,(f_1 * f_2)(z)$

3. Addition Theorem for Normal Distributions

Theorem: If independent random variables $X$ & $Y$ have normal distributions with the same mean $\mu$ and variance $\sigma^2$, then $Z = X + Y$ has mean $2\mu$ and variance $2\sigma^2$

## C. Multiplication or Division of Random Variables

1. Multiplication: $Z = XY$

   a. Convert $(x,y)$ to $(x,z)$: $x = x$ $\quad y = \frac{z}{x}$

   b. $J = \frac{1}{x}$

   c. $$g(z) = \int_{-\infty}^{\infty} f_1(x) f_2\left(\frac{z}{x}\right) \frac{dx}{|x|}$$ where $f(x,y) = f_1(x) f_2(y)$

II. C. (Continued)

2. Division: $Z = \frac{X}{Y}$

a. Convert $(x, y)$ to $(z, y)$: $x = yz$, $y = y$

b. $J = -y$

c. $$g(z) = \int_{-\infty}^{\infty} f_1(yz) f_2(y) \, |y| \, dy$$

D. The Gamma Distribution

1. For Gaussian distributed X, transform to $Y = X^2 \Rightarrow$ Gamma Distribution.

$$\boxed{g(y) = \begin{cases} 0 & y \le 0 \\ \dfrac{y^{\frac{1}{2}} e^{-y^2/2\sigma^2}}{(2\sigma^2)^{\frac{1}{2}} \sqrt{\pi}} & y > 0 \end{cases}}$$

Gamma Distribution for $p = \frac{1}{2}$, variance $\sigma^2$.

2. Def: General Gamma Distribution, $g(p, \sigma; y)$

$$\boxed{g(p, \sigma; y) = \begin{cases} 0 & y \le 0 \\ \dfrac{y^{p-1} e^{-y^2/2\sigma^2}}{(2\sigma^2)^p \Gamma(p)} & y > 0 \end{cases}}$$

3. Addition of Gamma Distributed Variables

a. Sum of gamma-distributed random variables $X_j$ with parameters $p_j$ but same $\sigma$ is gamma distributed with $\sigma$ and $p = \sum_j p_j$.

b. Ex: Two Variables, $X_1$ & $X_2$ with $g(p_1, \sigma; x_1)$ and $g(p_2, \sigma; x_2)$

i) $Y = X_1 + X_2 \Rightarrow g(p_1 + p_2, \sigma; y)$

4. Sum of squares of $n$ Gauss-distributed variables with common $\sigma^2$ yields a gamma distribution with $p = \frac{n}{2}$ and $\sigma$.

a. This arises when using method of lease squares to minimize deviations of measurements from a mean curve.

Hawes (5)

# III. Statistics

## A. Error Propagation

1. For $n$ repeated measurements $x_j$,

<u>Mean</u> $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ <u>Variance:</u> $\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\bar{x})^2$

2. Compute $\bar{f}$ and $\sigma^2(f)$ for function $f(x)$

a. $\bar{f} = \frac{1}{n}\sum_j f(x_j) = \frac{1}{n}\sum_j f(\bar{x}+e_j)$ where $e_j \equiv x_j - \bar{x}$

b. Taylor expanding $f(x)$ about $\bar{x}$, we can obtain

$$\boxed{\bar{f} = f(\bar{x}) + \frac{1}{2}\sigma^2 f''(\bar{x})}$$ Mean

c. What about spread of values $f(x_j)$?

i. Approximate $f_j = \bar{f} + f'(\bar{x})e_j$.

Thus, $\sigma^2(f) = \frac{1}{n}\sum_j (f_j-\bar{f})^2 = [f'(x)]^2\sigma^2$

ii. Therefore $\boxed{f(\bar{x}\pm\sigma) = f(\bar{x}) \pm f'(\bar{x})\sigma}$

3. <u>Multivariable Case:</u> $f(x,y)$ for $x_j, y_j$ measurements (independent)

$$\boxed{f(\bar{x}\pm\sigma_x, \bar{y}\pm\sigma_y) = f(\bar{x},\bar{y}) \pm \sqrt{f_x^2\sigma_x^2 + f_y^2\sigma_y^2}}$$

where $f_x \equiv \left.\frac{\partial f}{\partial x}\right|_{\bar{x},\bar{y}}$ $f_y = \left.\frac{\partial f}{\partial y}\right|_{\bar{x},\bar{y}}$

## B. Repeated Measurements

1. Consider making $n$ repeated measurements, each with variance $\sigma^2$.

a. Variance of the mean $\boxed{\sigma^2(\bar{x}) = \frac{\sigma^2}{n}}$ Individual measurement Variance.

## III. B. (Continued)

### 2. Sample Standard Deviation

a. The mean of measurements $x_j$, given by $\bar{x}$, will differ from true mean $\mu$ by some $\bar{x} = \mu + \alpha$

b. Difference from true mean $v_j = x_j - \mu$

Difference from arithmetic mean, $e_j = x_j - \bar{x}$

⇒ NOTE $v_j = e_j + \alpha$

c.
$$S^2 = \frac{1}{n}\sum_{j=1}^{n} v_j^2 = \frac{1}{n}\sum_{j=1}^{n} (e_j+\alpha)^2 = \frac{1}{n}\left[\sum_{j=1}^{n} e_j^2 + 2\alpha \cancelto{0}{\sum_{j=1}^{n} e_j} + \alpha^2 n\right]$$

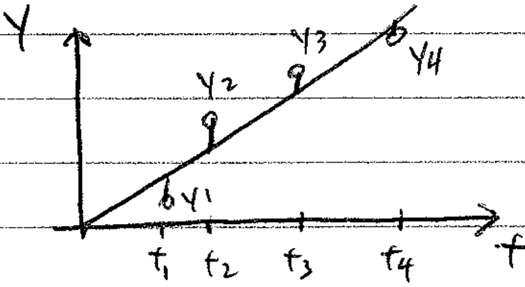
d. NOTE: Estimate of $\alpha^2 \approx \frac{S^2}{n}$ (repeated measurements), so
$$S^2\left(1-\frac{1}{n}\right) = \frac{1}{n}\sum_{j=1}^{\infty} e_j^2 = \sigma^2$$

e. Def Sample Standard Deviation
$$\boxed{s = \sqrt{\frac{\sum_j (x_j - \bar{x})^2}{(n-1)}}}$$

i. NOTE: $\frac{1}{n-1}$ factor accounts for possible error in $\bar{x}$ (Bessel's Correction)

## C. Curve Fitting

1. Consider sample measurements $y_j$ (with measurement error) taken at exact times $t_j$




2. Assume a known form $y = at$, where is $a$ must determined.
← regression coefficient

3a. $S = \sum_j (at_j - y_j)^2$

b. $\frac{dS}{da} = 2\sum_j t_j(at_j - y_j) = 0$

c. Solving for $a$: $\boxed{a = \frac{\sum_j t_j y_j}{\sum_j t_j^2}}$

Sample Standard Dev
$$s = \sqrt{\frac{\sum_j (y_j - at_j)^2}{(n-1)}}$$

## C. $\chi^2$ Distribution

1. When different measurements have different errors, we want a "weighted" least squares fit, or $\chi^2$ fit.

2. Def: Chi-Square, $\chi^2$

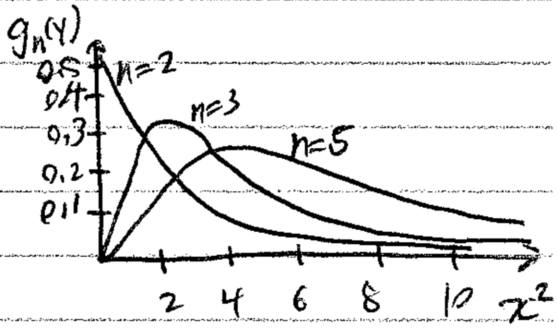
$$\boxed{\chi^2 = \sum_{j=1}^{n} \left( \frac{y_j - y(t_j, a, \ldots)}{\sigma_j} \right)^2}$$

where $y_j$ measurements have standard deviations $\sigma_j$

$a, \ldots$ are parameters of fit.

3. For normally distributed random variables, $\chi^2$ will yield a gamma distribution with $p = \frac{n}{2}$ and $\sigma = 1$.

$$g(\chi^2 = y) = \frac{y^{\frac{n}{2}-1} e^{-y/2}}{2^{n/2} \Gamma(\frac{n}{2})}$$




4. Typically, value of $\chi^2$ is comparable to number of data points

5. NOTE: If $\chi^2$ is based on $n$ data points and determination of $r$ parameters, number of degrees of freedom is $n-r$.

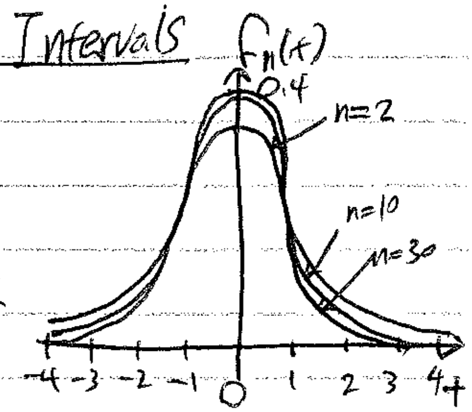
## D. Student t Distribution and Confidence Intervals

1. Def: Student t Distribution

a. $T = \frac{Y\sqrt{n}}{\sqrt{S/n}}$ where $Y = \frac{1}{n}\sum_{j=1}^{n} x_j - \mu$

$S = \sum_{j=1}^{n} x_j^2$

b. $\boxed{f_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}}$ ← Independent of variance of $x_j$



III. D. (Continued)

2a. <u>Confidence Interval</u>: Probability p that a measurement falls within

$x_0 - dx < x < x_0 + dx$

b. <u>Ex</u>: $x = 0.50 \pm 0.10$ with 90% confidence

c. For a Gauss distribution, the Student t Distribution provides a method to measure confidence intervals.

$$Y = \bar{X} - \mu = \frac{T\sqrt{S/n}}{\sqrt{n}}$$ where S is single value from data.

3. Probability $P(-c_p < t < +c_p) = p$

$$\Rightarrow P(-\infty < t < +c_p) = \boxed{\frac{1+p}{2} \equiv \hat{p}}$$ ← Used in table.

4. <u>Solve for $\mu$:</u>

$$\mu = \bar{X} - T\frac{\sqrt{S/n}}{\sqrt{n}} = \bar{x} - \frac{T\sigma}{\sqrt{n}}$$

Note: $\sigma = \sqrt{\frac{S}{n}}$ ← <u>Sample Standard Deviation!</u>

5. <u>Example</u>: <u>Confidence Interval</u>:

a. <u>Data</u>: 7.12, 4.95, 6.18, 5.69, 2.90, 8.47

b. Compute mean and 90% confidence interval:

c. Mean $= \frac{1}{n}\sum_j x_j = 5.885$

d. <u>Sample St. Dev</u>: $\sigma = \sqrt{\frac{\sum_j (x_j - \bar{x})^2}{n-1}} = 1.9035$

e. <u>90% Confidence</u>: $p = 0.9 \Rightarrow \hat{p} = \frac{1+0.9}{2} = 0.95$, <u>Degrees of Freedom</u> $n = 5$

<u>Table</u>: <u>Student t Dist</u>

| $\hat{p}$ | n=1 | n=2 | n=3 | n=4 | n=5 |
|---|---|---|---|---|---|
| 0.8 | 1.38 | 1.06 | 0.98 | 0.94 | 0.92 |
| 0.9 | 3.08 | 1.89 | 1.64 | 1.53 | 1.48 |
| 0.95 | 6.31 | 2.92 | 2.35 | 2.13 | 2.02 |

T = 2.02

$$\mu = \bar{x} \pm \frac{T\sigma}{\sqrt{n}} = 5.885 \pm \frac{(2.02)(1.9035)}{\sqrt{5}} = \boxed{5.885 \pm 1.720 \text{ (90\% confidence)}}$$