# MATH 545
# STATISTICAL INFERENCE &
# SAMPLING THEORY
# HANDOUT 5

Note: The basic building block of probability is set theory:
Suppose we have a well defined sample space S (a well defined universe U) and events $E_1$, $E_2$, etc. (sets $X_1$, $X_2$, etc.), yada, yada, yada. The basic ideas are grounded in the sets!

Moreover, much of the really powerful theorems and concepts are grounded in Set Theory, Real Analysis, and Topology. However, we have not studied Real Analysis to the extent (e.g.: no one has completed Math 351, 352 and I doubt many of us have completed Math 430) that we can ground fully all of the discussion of continuous random variables.

## DISCRETE RANDOM VARIABLES

**Definition 0:** If $\exists$ a non-negative function, $f$, defined from the domain[1] $(-\infty, \infty)$ to the codomain $(-\infty, \infty)$ meaning $f : \mathbb{R} \longrightarrow \mathbb{R}$ and if $\exists$ a set A which is the set of all $x \in \text{dom}(f)$ where $\Pr(X = x) > 0$ and that A has Lebsegue measurable $0$[2], then X is said to be a discrete random variable.

**Definition 1:** If X is a discrete random variable, the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is called the probability mass function (p. m. f.).

**Note 1**: Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be the probability mass function and A is the set of all $x \in \text{dom}(f)$ where $\Pr(X = x) > 0$. One can consider the function $f|_A : A \longrightarrow \mathbb{R}$ as the p.m.f. since $\forall\ x \in A^C$ $\Pr(X = x) = 0$. So, we have a condition that $f : \mathbb{R} \longrightarrow \mathbb{R}$ decomposes into

$f|_A : A \longrightarrow \mathbb{R}$ so that $f|_A\ (x) > 0\ \forall\ x \in A$ and

$f|_{A^C} : A^C \longrightarrow \mathbb{R}$ so that $f|_{A^C}(x) = 0\ \forall\ x \in A^C$

**Note 2**: So, without loss of generality we will say many times $f : A \longrightarrow \mathbb{R}$ is the probability mass function and A is the set $\text{dom}(f)$ so $\Pr(X = x) > 0$ for all $x \in A$ and not mention all the reals where $\Pr(X = x) = 0$.

---

[1] *Note: Most statistics texts call the domain of* $f(x)$ *the range…* don't ask me why, they just do; however, we will use the correct terminology in this class.

[2] Really all we need is for A to be finite or denumerable for our purposes. Therefore, $|A| \le \aleph_0$. Recall $|A| < \aleph_0$ means A is finite and Recall $|A| = \aleph_0$ means A is denumerable. So, $|A| \le \aleph_0$ means A is countable.

**<u>Theorem 1:</u>** A function serves as a p. m. f. of a discrete random variable iff its values $f(x)$ satisfy:

1. $f(x) \geq 0 \ \forall \ x \in \text{dom}(f)$

2. $\sum_{x} f(x) = 1$.

**<u>Definition 2</u>:** If X is a discrete random variable, the function given by $F(x) = \Pr(X \leq x)$ for each x in the domain of the function is called the probability distribution function or cumulative distribution function (c. d. f.)

$F(x) = \Pr(X \leq x) = \sum_{t \leq x} f(t) \quad \forall \ x \in (-\infty, \infty)$

**<u>Theorem 2</u>:** A function serves as a c. d. f. of a discrete random variable iff its values $F(x)$ satisfy:

1. $F(x) \longrightarrow 0$ as $x \longrightarrow -\infty$ or $\exists \ x_1 \in \mathbb{R} \ni F(x_1) = 0$

2. $F(x) \longrightarrow 1$ as $x \longrightarrow \infty$ or $\exists \ x_2 \in \mathbb{R} \ni F(x_2) = 1$

3. $F(x)$ is non-decreasing $\forall \ x \in (-\infty, \infty)$

**<u>Definition 3</u>:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the expected value (or mean) of X is $E[X] = \sum_{x} x \cdot f(x)$.

**<u>Notation</u>**: $E[X] = \mu = \mu_1'$

**<u>Definition 4</u>:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the $r^{th}$ moment about the origin of X is $E[X^r] = \sum_{x} x^r \cdot f(x)$.

**<u>Notation</u>**: $E[X^r] = \mu_r'$

**<u>Definition 5</u>:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the variance (or second moment about the mean) of X is $\text{Var}[X] = \sum_{x} (x - E[X])^2 \cdot f(x)$.

**<u>Notation</u>**: $\text{Var}[X] = \sigma^2 \quad \text{Var}[X] = E[(X - \mu)^2]$

**<u>Definition 6</u>:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the standard deviation of X is $SD[X] = \sqrt{\sum_{x} (x - E[X])^2 \cdot f(x)}$.

**<u>Notation</u>**: $SD[X] = \sigma$

**Definition 7:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the $r^{th}$ moment about the mean of X is $E[(X - \mu)^r] = \sum_x (x - E[X])^r \cdot f(x)$.

**Notation**: $E[(X - \mu)^r] = \mu_r$

**Theorem 3:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then $\text{Var}[X] = \mu_2' - \mu^2 = E[X^2] - (E[X])^2$

**Definition 8:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the coefficient of skewness, $\eta_3$, is $\frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\sqrt{\mu_2})^3}$

**Definition 9:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x, then the coefficient of kurtosis, $\eta_4$, is $\frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^2}$

**Theorem 4:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and $g(X)$ is a function of X , then the expected value (or mean) of $g(X)$ is $E[g(X)] = \sum_x g(X) \cdot f(x)$.

**Theorem 5:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and T is a linear transformation of X

(i.e.: $T = \alpha X + \beta) \ni \alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, then the expected value (or mean) of T is

$E[T] = E[\alpha X + \beta] = \alpha E[X] + \beta$.

**Corollary 5.1:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and T is a zero linear transformation of X (i.e.: $T = \alpha X + 0$) then the expected value (or mean) of T is $E[T] = E[\alpha X] = \alpha E[X]$.

**Corollary 5.2:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and T is a constant (i.e.: $T = \beta \ni \beta \in \mathbb{R}$ then the expected value (or mean) of T is $E[T] = E[\beta] = \beta$.

**Theorem 6:** If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and T is a linear transformation of X (i.e.: $T = \alpha X + \beta) \ni \alpha \in \mathbb{R}$,

$\beta \in \mathbb{R}$, then the variance of T is $E[(T - \mu_T)^2] = \text{Var}[\alpha X + \beta] = \alpha^2 \text{Var}[X]$.

**<u>Theorem 7 (Chebyshev's Theorem or Tchebyshev's Theorem)</u>:**
If X is a discrete random variable and the function given by $f(x) = \Pr(X = x)$ for each x in the domain of the function is the p. m. f. at x and $\mu$ and $\sigma$ are the mean and standard deviation of X, then for any $k > 0$ $(k \in \mathbb{R})$ the probability is *at least* $1 - \frac{1}{k^2}$ that X will take on a value within k standard deviations of the mean.

i.e.: $\Pr(|X - \mu| < k \cdot \sigma) \geq 1 - \frac{1}{k^2}$

# RANDOM VARIABLE (DISCRETE) P. M. F. S AND MOMENTS OF IMPORT.

**The Bernoulli** trial is a probabilistic (or stochastic) experiment that can have one of two outcomes, success (X = 1) or failure (X = 0) in which the probability of success is p. The parametre is p. We call the variable X a Bernoulli (discrete) random variable.

$p \in (0, 1)$

$x \in \{0, 1\}$

$$\text{Ber}(x, p) = \Pr(X = x) = \begin{cases} p, & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{else} \end{cases}$$

$\mu = p$

$\mu_r' = p \quad \forall\, r \in \mathbb{N}$

$\sigma^2 = p(1 - p)$

**The Binomial** variate is the number (x) of successes in n-independent Bernoulli trials where the probability of success at each trial is p. The parametres are p and n (the number of trials).

We call the variable X a binomial random variable.

$p \in (0, 1)$

$n \in \mathbb{N}$

$x \in \{0, 1, 2, \ldots, (n - 1), n\}$

$$\text{Bin}(x, p, n) = \Pr(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{else} \end{cases}$$

$\mu = n \cdot p$

$\mu_2' = n \cdot p \cdot (n \cdot p + (1 - p))$

$\mu_3' = n \cdot p((n-1) \cdot (n-2) \cdot p^2 + 3 \cdot p \cdot (n - 1) + 1)$

$\sigma^2 = n \cdot p \cdot (1 - p)$

$\mu_3 = n \cdot p \cdot (1 - p) \cdot ((1 - p) - p)$

$\mu_4 = n \cdot p \cdot ((1 + 3p \cdot (1-p) \cdot (n - 2))$

**The Poisson** variate (x) is as follows. The parametre is $\lambda$ (a positive real number). We call the variable X a Poisson random variable.

$\lambda \in (0,\infty) \quad x \in \mathbb{N}^*$

$$\text{Pois}(x, \lambda) = \Pr(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & x = 0,1,2,\ldots \\ 0 & \text{else} \end{cases}$$

$\mu = \lambda$ $\quad \mu_2' = \lambda + \lambda^2$ $\quad \mu_3' = \lambda[(\lambda + 1)^2 + \lambda]$ $\quad \mu_4' = \lambda[\lambda^3 + 6\lambda^2 + 7\lambda + 1]$

$\sigma^2 = \lambda$ $\quad \mu_3 = \lambda$ $\quad \mu_4 = \lambda(1 + 3\lambda)$ $\quad \mu_5 = \lambda(1 + 10\lambda)$ $\quad \mu_6 = \lambda(1 + 25\lambda + 15\lambda^2)$

# CONTINUOUS RANDOM VARIABLES

**<u>Definition 1:</u>** If $\exists$ a non-negative function, *f*, defined $\forall$ $x \in (-\infty, \infty)$ such that for any measurable[3] set of positive real measure $A \subseteq \mathbb{R}$, $\Pr(X \in A) = \displaystyle\int_{x \in A} f(x)dx$, then X is said to be a continuous random variable.

**<u>Definition 2:</u>** If X is a continuous random variable such that the function given by *f*(x) from definition 1 exists for each x in the domain of the function, then *f*(x) is called the probability density function (p. d. f.).

*Note: Recall most statistics texts call the domain of* *f*(x) *the range…* don't ask me why, they just do; however, we will use the correct terminology in this class.

**<u>Theorem 1</u>**: If X is a continuous random variable such that the function given by *f*(x) from definition 1 exists for each x in the domain of the function, then Pr (X = a) is 0.

**<u>Theorem 2</u>**: If X is a continuous random variable such that the function given by *f*(x) from definition 1 exists for each x in the domain of the function, and A is a set such that |A| = n ($n \in \mathbb{N}$) then $\Pr(X \in A) = \displaystyle\sum_{x \in A} (\Pr(X = x))$ is 0.

**<u>Theorem 3[4]</u>**: If X is a continuous random variable such that the function given by *f*(x) from definition 1 exists for each x in the domain of the function, and A is a set such that $|A| = \aleph_0$ then $\Pr(X \in A) = \displaystyle\sum_{x \in A} (\Pr(X = x))$ is 0.

---

[3] This is where we get into a problem. In order to really understand this one needs to take Math 463, Math 485, and perhaps even more. Suffice it to say that we will only consider sets of positive measure in this class.

[4] We will *not* prove this theorem in this class. More prerequisite knowledge and experience is necessary.

**<u>Theorem 4</u>**[5]: If X is a continuous random variable such that the function given by *f*(x) from definition 1 exists for each x in the domain of the function, and A is a set such that its measure is zero then Pr $(X \in A) = \sum_{x \in A} (\Pr(X = x))$ is 0.

**<u>Theorem 5:</u>** A function serves as a p. d. f. of a continuous random variable iff its values *f*(x) satisfy:

1. $f(x) \geq 0 \ \forall \ x \in (-\infty, \infty)$
2. $\int_{x} f(x)dx = 1$.

**<u>Definition 3</u>:** If X is a continuous random variable and the function *f*(x) is the p. d. f. of X, then

$$\Pr(a < X < b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \Pr(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

**<u>Definition 4</u>:** If X is a continuous random variable, the function given by *F*(x) = Pr(X ≤ x) for each x in the domain of the function is called the probability distribution function or cumulative distribution function (c. d. f.)

$$F(x) = \Pr(X \leq x) = \Pr(X < x) = \int_{t<x} f(t)dt = \int_{t \leq x} f(t)dt = = \int_{-\infty}^{x} f(x)dx \quad \forall \ x \in (-\infty, \infty)$$

**<u>Theorem 6</u>:** A function serves as a c. d. f. of a continuous random variable iff its values *F*(x) satisfy:

1. $F(x) \longrightarrow 0$ as $x \longrightarrow -\infty$ or $\exists \ x_1 \in \mathbb{R} \ni F(x_1) = 0$

2. $F(x) \longrightarrow 1$ as $x \longrightarrow \infty$ or $\exists \ x_2 \in \mathbb{R} \ni F(x_2) = 1$

3. *F*(x) is non-decreasing $\forall \ x \in (-\infty, \infty)$

**<u>Definition 4</u>:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the expected value (or mean) of X is $E[X] = \int_{x} x \cdot f(x)dx = \int_{-\infty}^{\infty} x \cdot f(x)dx$

$E[X\} = \mu = \mu_1'$

**<u>Definition 5</u>:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the r^th^ moment about the origin of X is $E[X^r] = \int_{x} x^r \cdot f(x)dx = \int_{-\infty}^{\infty} x^r \cdot f(x)dx$

. $E[X^r] = \mu_r'$

---

[5] We will *not* prove this theorem in this class. More prerequisite knowledge and experience is necessary.

**Definition 6:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the variance (or second moment about the mean) of X is Var[X] =

$$\int_{x} (x - E[X])^2 \cdot f(x)dx = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f(x)dx \quad .$$

Var[X] = $\sigma^2$ Var[X] = $E[(X - \mu)^2]$

**Definition 7:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the standard deviation of X is $SD[X] = \sqrt{\int_{-\infty}^{\infty} (x - E[X])^2 \cdot f(x)dx}$ . SD[X] = $\sigma$

**Definition 8:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the r^th^ moment about the mean of X is $E[(X - \mu)^r] = \int_{-\infty}^{\infty} (x - E[X])^r \cdot f(x)dx$ .

$E[(X - \mu)^r] = \mu_r$

**Theorem 7:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then Var[X] = $\mu_2' - \mu^2 = E[X^2] - \left(E[X]\right)^2$

**Definition 9:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the coefficient of skewness, $\eta_3$ , is $\frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\sqrt{\mu_2})^3}$

**Definition 10:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then the coefficient of kurtosis, $\eta_4$ , is $\frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^{\,2}}$

**Theorem 8:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and *g*(X) is a function of X , then the expected value (or mean) of *g*(X) is E[*g*(X)] =

$$\int_{-\infty}^{\infty} (g(x)) \cdot f(x)dx .$$

**Theorem 9:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and T is a linear transformation of X (i.e.: T = $\alpha X + \beta$) ∋

$\alpha \in \mathbb{R}$ , $\beta \in \mathbb{R}$, then the expected value (or mean) of T is E[T] = E[(i.e.: T = $\alpha X + \beta$)] = $\alpha E[X] + \beta$.

**Corollary 9.1:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and T is a zero linear transformation of X

(i.e.: $T = \alpha X \ni$ (i.e.: $T = \tilde{\alpha} X + 0$) $\in \mathbb{R}$ then the expected value (or mean) of T is $E[T] = E[\alpha X] = \alpha E[X]$.

**Corollary 9.2:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and T is a constant (i.e.: $T = \beta \ni \beta \in \mathbb{R}$ then the expected value (or mean) of T is $E[T] = E[\beta] = \beta$.

**Theorem 10:** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and T is a linear transformation of X (i.e.: $T = \alpha X + \beta$) $\ni$ $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, then the variance of T is $E[(T - \mu_T)^2] = Var[\alpha X + \beta] = \alpha^2 Var[X]$.

**Lemma 11.1 (Markov's Inequality):** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x, then for any $k > 0$ ($k \in \mathbb{R}$) the probability that X is greater than or equal to k is less than or equal to the mean divided by k.

$$Pr(X \geq k) \leq \frac{E[X]}{k}$$

**Theorem 11 (Chebyshev's Theorem or Tchebyshev's Theorem):** If X is a continuous random variable and the function given by *f*(x) for each x in the domain of the function is the p. d. f. at x and $\mu$ and $\sigma$ are the mean and standard deviation of X, then for any $k > 0$ ($k \in \mathbb{R}$) the probability is *at least* $1 - \frac{1}{k^2}$ that X will take on a value within k standard deviations of the mean.

$$Pr(|X - \mu| < k \cdot \sigma) \geq 1 - \frac{1}{k^2}$$

# RANDOM VARIABLE (CONTINUOUS) P. M. F. S AND MOMENTS OF IMPORT.

**The Uniform** random variable is a probabilistic (or stochastic) experiment that can have any outcome on an interval of $\mathbb{R}$ such that it can be β (X = 1), α (X = 0), or any value between such the probability is one divided by β - α for any of these values; the probability is zero otherwise. An application of the uniform random variable is the round off difference between recorded and true values of physical quantities. It provides (oft times) a fair approximation over a relatively narrow range for a random variable whose distribution is not uniform.

$x \in [\alpha, \beta]$

$$\text{Uni}(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & x \in [\alpha, \beta] \\ 0 & \text{else} \end{cases}$$

$$\mu = \frac{\beta + \alpha}{2}$$

$$\mu_r' = \frac{\beta^{r+1} - \alpha^{r+1}}{(\beta - \alpha)(r+1)} \qquad \mu_r = \begin{cases} 0 & r \in O \\ \dfrac{(\beta - \alpha)^r}{2^r (r+1)} & r \in E \end{cases}$$

where O is the set of odd natural numbers and E is the set of even natural numbers

$$\sigma^2 = \frac{(\beta - \alpha)^2}{12} \qquad \eta_3 = 0 \qquad \eta_4 = 1.8$$

**The Normal[6] (Gaussian[7])** random variable is a probabilistic (or stochastic) experiment that can have any outcome on $\mathbb{R}$. The parametres are $\mu$ and $\sigma^2$ (or $\mu$ and $\sigma$). Thus, it is defined by its mean and variance. Its applications are many and its use *quite* important. A substantial number of empirical studies have indicated that the normal function provides an adequate representation of, or at least a decent approximation to, the distributions of many physical, mental, economic, biological, and social variables. For example: meteorological data (temperature and rainfall), measurements of living organisms (height or weight of humans, populations, etc.), scores on aptitude tests, physical measurements of of manufactured parts, instrumental errors, deviations from social norms, and above all else as we take means of means the limiting distribution is normal.

$x \in \mathbb{R}$, $\mu \in \mathbb{R}$, and $\sigma \in \mathbb{R}^+ (0, \infty)$

$$\text{Nor}(x, \mu, \sigma^2) = \frac{e^{\frac{-(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \quad \forall x \in (-\infty, \infty) \qquad \text{Nor}(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\cdot\sigma\cdot e^{\frac{(x-\mu)^2}{2\sigma^2}}}$$

$\mu = \mu$ median = $\mu$ mode = $\mu$

$$\mu_r = \begin{cases} 0 & r \in O \\ \dfrac{\sigma^r \cdot r!}{2^{r/2}(\frac{r}{2})!} & r \in E \end{cases}$$

where O is the set of odd natural numbers and E is the set of even natural numbers

$\sigma^2 = \sigma^2$ $\eta_3 = 0$ $\eta_4 = 3$

**A special case of the normal family is the standard normal function:**

$$\text{Nor}(z, 0, 1) = \frac{e^{\frac{-z^2}{2}}}{\sqrt{2\pi}} \quad \forall z \in (-\infty, \infty)$$

*Note:* we will use Z ~ Nor (z, 0, 1).

where Z = $Z = \dfrac{X - \mu_X}{\sigma_X}$

[6] The *most* important of the continuous distributions.

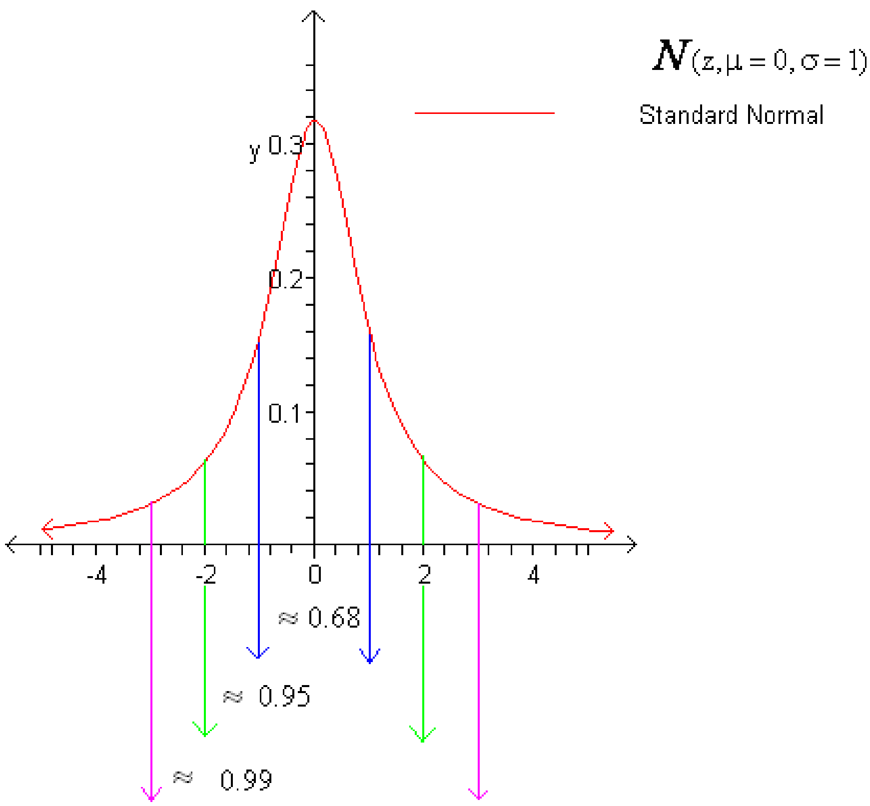
[7] Gaussian after K. F. Gauss, "the Prince of Mathematics," who laid much of the ground work for Modern Probability & Statistics (1809 and later). The Normal curve family of functions was first considered by DeMoivre (1733). Modern Probability & Statistics, however, was formalised by R. A. Fisher & K. Pearson.

The c. d. f. of the standard normal function[8] (also called the unidimensional normal ogive function) is:

$$\text{CDFN}\ (z, 0, 1\ ) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi} \cdot e^{t^2/2}}\, dt \quad \forall\ z \in (-\infty, \infty)$$

**Theorem 12 (DeMoivre - Laplace) :** Let X ~ Bin (x, n, p). Let $Y = \frac{X - np}{\sqrt{np(1-p)}}$.

As $n \longrightarrow \infty$, it is the case that $Y \longrightarrow Z$.




[8] Note it is a non-integrable [closed form] function.

**The Birnbaum[9] (logistic)** random variable is a probabilistic (or stochastic) experiment that can have any outcome on $\mathbb{R}$.. It was created to approximate the normal ogive model for dichotomous response vectors. The parametres are θ, a, b, and c. Expressing the normal ogive curve relative to a response vector correct as:

$$P(x_j = 1 \mid a_i,\ \ b_i,\ \ c_i,\ \ \theta_j) = c_i + (1 - c_i) \int_{-\infty}^{a_i(\theta_j - b_i)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

Thus, the Birnbaum the unidimensional logistic models are specified:

$$\text{Bir}\ (x, a, b, c, \theta) = P(x_j = 1 \mid a_i,\ \ b_i,\ \ c_i,\ \ \theta_j) = c_i + (1 - c_i) \frac{e^{Da_i(\theta_j - b_i)}}{1 + e^{Da_i(\theta_j - b_i)}}\ ,\ \ i \in \mathbb{N}$$

$P(x_j = 1 \mid a_i,\ \ b_i,\ \ c_i,\ \ \theta_j)$ is the probability that a randomly chosen $j^{th}$ thing with parametre $\theta_j$ has response i correct [X = 1], D is a scaling constant, $a_i$ is the item discrimination parametre, $b_i$ is the item difficulty parametre, and $c_i$ is the item pseudo-guessing parametre. $a_i \in (0, \infty)$. $b_i \in (-\infty, \infty)$. $c_i \in [0,1]$. Typically, D = 1.7 or 1.702. If $a_i$ = 0 and $c_i$ =0, then the model is referred to as the unidimensional one parametre logistic model (UIRT1PL). If $c_i$ =0, then the model is referred to as the unidimensional two parametre logistic model (UIRT2PL). Thus, if $a_i$, bi, and $c_i$ are not fixed, then the model is referred to as the unidimensional three parametre logistic model (UIRT3PL).

Further, suppose Bir(1.702x) denotes a logistic continuous density function and Nor(x) denotes a normal ogive continuous density function (which is the normal cumulative distribution function), Haley (1952) argued that | Nor(x) - Bir(1.702x) | < .01 for $x \in (-\infty,\ \infty)$.

[9] I am including this function for informational purposes. It is interesting to note that the normal ogive is a difficult function to work with; thus, the need for an approximation to the normal ogive. Further, this should help you infer that much of the work in Probability and Statistics involves modeling.

**The Chi-Square (or chi-squared)** random variable is a probabilistic (or stochastic) experiment that can have any outcome on $(0, \infty)$. The parametre is $\nu$ (nu). It is a special case of the Gamma variate (you will prove this later). The parametre, $\nu$, is called the degrees of freedom (referenced quite frequently in applied probability and statistics).

It plays a significant role in inferential statistics and sampling theory, especially with regard to inferences about variances[10]. It also is highly useful in non-parametric statistics (for example, in tests of the lack of fit of data and frequency problems).

$$\text{Chi}(x, \nu) = \begin{cases} \dfrac{x^{(\nu-2)/2}}{2^{\nu/2}\Gamma(\nu/2)e^{(x/2)}} & x \in (0, \infty) \\ 0 & \text{else} \end{cases}$$

Notation: $\text{Chi}(x, \nu) = \chi^2_\nu$

$\mu = \nu$ $\qquad \sigma^2 = 2\nu$

$\eta_3 = \dfrac{4}{\sqrt{2\nu}}$ $\qquad \eta_4 = 3(1 + \dfrac{4}{\nu})$

mode = $\nu - 2 \ni \nu \geq 2$

median $\approx \nu - \dfrac{3}{2}$ when $\nu$ is 'large'

$$\mu_r' = 2^r \prod_{i=0}^{r-1} [i + (\frac{\nu}{2})] = \frac{2^r \cdot \Gamma(r + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}$$

---

[10] Which is discussed (for proofs, etc.) in Math 301-302-403-540.
. It turns out that if $X \sim \text{Nor}(x, \mu, \sigma)$, then $X^2 = Y \sim \text{Chi}(y, 1)$.

**The Student (or t or Gosset)** random variable is a probabilistic (or stochastic) experiment that can have any outcome on $(-\infty, \infty)$. The parametre is $\nu$ such that $\nu \in \mathbb{N}$. The parametre, $\nu$, is called the degrees of freedom (referenced quite frequently in applied probability and statistics).

It plays a significant role in inferential statistics and sampling theory, especially with regard to inferences about differences of the means between two groups[11] under certain conditions (which usually are not verified by the social scientists using it - aargh).

$$\text{Stu}(x, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\frac{\nu}{2})[1+\frac{x^2}{\nu}]^{(\nu+1)/2}} \quad x \in (-\infty, \infty). \quad \text{Notation:} \quad \text{Stu}(x, \nu) = t_\nu$$

$\mu = 0$ when $\nu > 1$ $\quad \sigma^2 = \dfrac{\nu}{\nu - 2}$ when $\nu > 2$ $\quad$ mode $= 0$ $\quad$ median $= \alpha$

$$\mu_r = \begin{cases} 0 & r \in O \\ AM & r \in E \end{cases} \qquad \eta_3 = 0 \text{ when } \nu > 3 \qquad \eta_4 = \frac{3(\nu-2)}{\nu-4} \text{ when } \nu > 4$$

Note: AM denotes "a mess" in the definition of $\mu_r$.

---

[11] Which is discussed (for proofs, etc.) in Math 301-302-403-540.

**F (variance ratio) (or Fisher - Snedecor)** random variable is a probabilistic (or stochastic) experiment that can have any outcome on $[0, \infty)$. The parametres are $\nu$ and $\omega$ (omega) such that $\nu \in \mathbb{N}$ and $\omega \in \mathbb{N}$. The parametres, $\nu$ and $\omega$, are both called the degrees of freedom (referenced quite frequently in applied probability and statistics).

It plays a significant role in inferential statistics and sampling theory, especially with regard to inferences about differences of the means between multiple groups[12] under certain conditions (which usually are not verified by the social scientists using it - aargh).

$$\text{Fish}(x, \nu, \omega) = \frac{\Gamma(\frac{\nu+\omega}{2})(\frac{\nu}{\omega})^{\nu/2} x^{(\nu-2)/2}}{\Gamma(\frac{\nu}{2})\Gamma(\frac{\omega}{2})[1+\frac{\nu x}{\omega}]^{(\nu+\omega)/2}} \quad x \in [0, \infty).$$

Notation: $\text{Fish}(x, \nu) = F_{\nu,\omega}$

$\mu = \dfrac{\omega}{\omega+2}$ when $\omega > 2$ $\qquad \sigma^2 = \dfrac{2\omega^2(\nu+\omega-2)}{\nu(\omega-2)^2(\omega-4)}$ when $\omega > 4$

**Last revised 22 Oct. 2007** © 2000 – 2007, M. P. M. M. M.

[12] Which is discussed (for proofs, etc.) in Math 301-302-403-540.