

**tags:**    distributionrelationships dist

**NOTE:**

**field:**    <sub>54</sub>

**field:**     $X \sim Gamma(a, b)$   $P(X \leq X) =$

**field:**     $X \sim Gamma(a, b)$   $P(X \leq X) = P(Y \geq a)$  Where  $Y \sim Pois(x/b)$

**NOTE:**

**field:**    <sub>55</sub>

**field:**

$$X_1, \dots, X_n \sim iidN(0, 1)$$
$$\sum X_i \overset{?}{\sim}$$

**field:**

$$X_1, \dots, X_n \sim iidN(0, 1)$$
$$\sum X_i \sim N(0, n)$$

**NOTE:**

**field:**    <sub>56</sub>

**field:**

$$X_1, \dots, X_n \sim iidN(\mu_i, \sigma_i^2)$$
$$\sum X_i \overset{?}{\sim}$$

**field:**

$$X_1, \dots, X_n \sim iid N(\mu_i, \sigma_i^2)$$
$$\sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$$

**NOTE:**

**field:** <sub>57</sub>

**field:**

$$X \sim N(\mu, \sigma^2)$$
$$aX + b \stackrel{?}{\sim}$$

**field:**

$$aX + Y \sim N(a\mu + b, a^2\sigma^2)$$

**NOTE:**

**field:** <sub>58</sub>

**field:**  $X \sim Binom(1, p) \stackrel{?}{\sim}$

**field:**  $X \sim Bern(p)$

**NOTE:**

**field:** <sub>59</sub>

**field:**  $X \sim NegBinom(1, p) \stackrel{?}{\sim}$

**field:**  $X \sim Geom(p)$

**NOTE:**

**field:** <sub>60</sub>

**field:**  $X \sim \text{Gamma}(1, \theta) \overset{?}{\sim}$

**field:**  $X \sim \text{Exp}(\theta)$

**NOTE:**

**field:** <sub>61</sub>

**field:**  $X \sim \text{Exp}(\theta) \overset{?}{\sim}$

**field:**  $X \sim \text{Gamma}(1, \theta)$

**NOTE:**

**field:** <sub>62</sub>

**field:**  $X \sim \text{Gamma}(v/2, 1/2) \overset{?}{\sim}$

**field:**  $X \sim \chi^2(v)$

**NOTE:**

**field:** <sub>63</sub>

**field:**

$$X \sim \chi^2(v) \overset{?}{\sim}$$

**field:**

$$X \sim \text{Gamma}(v/2, 2)$$

**NOTE:**

**field:** <sub>64</sub>

**field:**

$$X \sim \chi^2(2) \stackrel{?}{\sim}$$

**field:**

$$X \sim \exp(2)$$

**NOTE:**

**field:** <sub>65</sub>

**field:**

$$X \sim Weibull(1, \beta) \stackrel{?}{\sim}$$

**field:**

$$X \sim \text{Exp}(\beta)$$

**NOTE:**

**field:** <sub>66</sub>

**field:**  $X_1, X_2 \sim \chi^2(v_i)$  independent  $\frac{X_1/v_1}{X_2/v_2}$

**field:**

$$\frac{(X_1/v_1)}{(X_2/v_2)} \sim F(v_1, v_2)$$

**NOTE:**

**field:** <sup>67</sup>

**field:**

$$X \sim beta(1, 1) \stackrel{?}{\sim}$$

**field:**

$$X \sim Unif(0, 1)$$

**NOTE:**

**field:** <sup>68</sup>

**field:**

$$X \sim Unif(0, 1) \stackrel{?}{\sim}$$

**field:**

$$X \sim beta(1, 1)$$

**NOTE:**

**field:** <sup>69</sup>

**field:** Special case of t

$$X \sim t_1 \stackrel{?}{\sim}$$

**field:**

$$X \sim Cauchy(0, 1)$$

**NOTE:**

**field:** <sub>70</sub>

**field:** Scaled Gamma

$$X \sim Gamma(\alpha, \beta), Y = aX \stackrel{?}{\sim}$$

**field:**

$$Y \sim Gamma(\alpha, a\beta)$$

**NOTE:**

**field:** <sub>71</sub>

**field:** Scaled Exponential

$$X \sim Exp(\lambda), Y = aX \stackrel{?}{\sim}$$

**field:**

$$Y \sim Exp(a\lambda)$$

**NOTE:**

**field:** <sub>72</sub>

**field:** Sum of Exponential, equal rate  $X_i \sim \text{Exp}(\lambda), Y = \sum X_i$

**field:**

$$Y \sim \text{Gamma}(n, \lambda)$$

**NOTE:**

**field:** <sub>73</sub>

**field:**

$$X \sim \text{Exp}(\lambda), Y = e^{-x}$$

**field:**

$$Y \sim \text{Beta}(\lambda, 1)$$

**NOTE:**

**field:** <sub>74</sub>

**field:** Min of Exponential

$$X_1, \dots, X_n \text{Exp}(\lambda_i), Y = \min(X_i) \stackrel{?}{\sim}$$

**field:**  $Y \sim \text{exp}(\sum \lambda_i)$

**NOTE:**

**field:** <sub>75</sub>

**field:** Min of Uniform

$$X_i \sim Unif(0, 1), Y = \lim n \min(X_i) \stackrel{?}{\sim}$$

**field:**

$$Y \sim Exp(1)$$

**NOTE:**

**field:** <sub>76</sub>

**field:**

$$X \sim Beta(\alpha, \beta), Y = (1 - X)$$

**field:**

$$Y \sim Beta(\beta, \alpha)$$

**NOTE:**

**field:** <sub>77</sub>

**field:**  $X \sim F_X(X), Y = F_X^{-1}(X)$

**field:**  $Y \sim Unif(0, 1)$

**NOTE:**

**field:** <sub>78</sub>

**field:**  $X \sim N(\mu, \sigma^2), Y = e^X$



**field:**  $Y \sim \text{lognormal}(\mu, \sigma^2)$

**NOTE:**

**field:** <sub>79</sub>

**field:**  $X \sim \exp(\beta), Y = X^{1/z}$

**field:**  $Y \sim \text{Weibull}(z, \beta)$

**NOTE:**

**field:** <sub>80</sub>

**field:** Square of Normal  $X \sim N(0, 1), Y = X^2$

**field:**  $Y \sim \chi^2(1)$

**NOTE:**

**field:** <sub>81</sub>

**field:** Square of t  $X \sim t(v), Y = X^2$

**field:**  $Y \sim F(1, v)$

**NOTE:**

**field:** <sub>82</sub>

**field:** Sum of Poisson  $X_i \sim \text{Poisson}(\mu_i) Y = \sum X_i$

**field:**  $Y \sim \text{Poisson}(\sum \mu_i)$

**NOTE:**

**field:** <sub>83</sub>

**field:** Sum of Gamma  $X_i \sim \text{Gamma}(\alpha_i, \beta), Y = \sum X_i$

**field:**  $Y \sim \text{Gamma}(\sum \alpha_i, \beta)$

**NOTE:**

**field:** <sub>84</sub>

**field:** Sum of independent Chi-squared  $X_i \sim \chi^2(v_i) Y = \sum X_i$

**field:**  $Y \sim \chi^2(\sum v_i)$

**NOTE:**

**field:** <sub>85</sub>

**field:**  $X, Y$  independent  $X, Y \sim N(0, 1), X/Y$

**field:**  $X/Y \sim \text{Cauchy}(0, 1)$

**NOTE:**

**field:** <sub>86</sub>

**field:**  $X_1, X_2 \sim \text{gamma}(\alpha_i, 1)$  independent,  $\frac{X_1}{X_1 + X_2}$

**field:**

$$\frac{X_1}{X_1 + X_2} \sim \text{beta}(\alpha_1, \alpha_2)$$

**NOTE:**

**field:** <sub>87</sub>

**field:**  $X_1, X_2 \sim \text{gamma}(\alpha_i, \beta_i)$  independent,  $\frac{\beta_2 X_1}{\beta_2 X_1 + \beta_1 X_2}$

**field:**

$$\frac{\beta_2 X_1}{\beta_2 X_1 + \beta_1 X_2} \sim \text{beta}(\alpha_1, \alpha_2)$$

**NOTE:**

**field:** <sub>88</sub>

**field:**  $X, Y$  independent  $\exp(\mu)$   $X - Y$

**field:**  $X - Y \sim \text{double exponential}(0, \mu)$

**NOTE:**

**field:** <sub>89</sub>

**field:**  $X \sim \text{Gamma}(\alpha, \beta)$   $Y = 1/X$

**field:** Inverted Gamma

**NOTE:**

**field:** <sub>90</sub>

**field:** Bernoulli( $p$ ),  $E(X) =$ ,  $V(X) =$

**field:** Bernoulli( $p$ ),  $E(X) = p$ ,  $V(X) = p(1 - p)$

**NOTE:**

**field:** <sub>91</sub>

**field:** Discrete Uniform  $N$ ,  $E(X) =$ ,  $V(X) =$

**field:** Discrete Uniform  $N, E(X) = \frac{N+1}{2}, V(X) = \frac{(N+1)(N-1)}{12}$

**NOTE:**

**field:** <sub>92</sub>

**field:** Cauchy( $\theta, \sigma$ ),  $E(X) =, V(X) =$

**field:** Cauchy( $\theta, \sigma$ ),  $E(X) = na, V(X) = na$

**NOTE:**

**field:** <sub>93</sub>

**field:** Double Exponential( $\mu, \sigma$ ),  $E(X) =, V(X) =$

**field:** Double Exponential( $\mu, \sigma$ ),  $E(X) = \mu, V(X) = 2\sigma^2$

**NOTE:**

**field:** <sub>94</sub>

**field:**  $F(v_1, v_2), E(X) =, V(X) =$

**field:**  $F(v_1, v_2), E(X) = \frac{v_1}{v_2-2}, V(X) = 2\left(\frac{v_2}{v_2-2}\right)^2 \frac{(v_1+v_2-2)}{v_1(v_2-4)}$

**NOTE:**

**field:** <sub>95</sub>

**field:** Mean and Variance for Distributions not on bible (but in CB)

- Double Exponential( $\mu, \sigma$ ),  $E(X) =$ ,  $V(X) =$
- $F(v_1, v_2)$ ,  $E(X) =$ ,  $V(X) =$
- Logistic( $\mu, \beta$ ),  $E(X) =$ ,  $V(X) =$
- Lognormal( $\mu, \sigma^2$ ),  $E(X) =$ ,  $V(X) =$
- Pareto( $\alpha, \beta$ ),  $E(X) =$ ,  $V(X) =$
- $t(v)$ ,  $E(X) =$ ,  $V(X) =$
- Weibull( $\gamma, \beta$ ),  $E(X) =$ ,  $V(X) =$

**field:** Mean and Variance. for Distributions not on bible (but in CB)

- Logistic( $\mu, \beta$ ),  $E(X) = \mu$ ,  $V(X) = \frac{\pi^2 \beta^2}{3}$
- Lognormal( $\mu, \sigma^2$ ),  $E(X) = e^{\mu + (\sigma^2/2)}$ ,  $V(X) = e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2}$
- Pareto( $\alpha, \beta$ ),  $E(X) = \frac{\beta \alpha}{\beta - 1}$ ,  $V(X) = \frac{\beta \alpha^2}{(\beta - 1)^2 (\beta - 2)}$
- $t(v)$ ,  $E(X) = 0$ ,  $V(X) = \frac{v}{v - 2}$
- Weibull( $\gamma, \beta$ ),  $E(X) = \beta^{1/\gamma} \Gamma(1 + 1/\gamma)$ ,  $V(X) = \beta^{2/\gamma} (\Gamma(1 + 2/\gamma) - \Gamma^2(1 + 1/\gamma))$

**tags:** Calculus calc

**NOTE:**

**field:** <sub>96</sub>

**field:**  $\int_0^\infty e^{-x^2/2} =$

**field:**  $\int_0^\infty e^{-x^2/2} = \sqrt{\pi/2}$

**NOTE:**

**field:** <sub>97</sub>

**field:**  $\int_0^\infty x^{a-1} e^{-x/b} =$

**field:**  $\int_0^\infty x^{a-1} e^{-x/b} = \Gamma(a) b^a$

**NOTE:**

**field:** <sub>98</sub>

**field:**  $\int_0^1 x^{a-1} (1-x)^{b-1} =$

**field:**  $\int_0^1 x^{a-1} (1-x)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

**NOTE:**

**field:** <sub>99</sub>

**field:**  $\log(x) = y, x =$

**field:**  $\log(x) = y, x = e^y$

**NOTE:**

**field:** <sub>100</sub>

**field:**  $\lim_{x \rightarrow \infty} (1 + \frac{a}{x})^x =$

**field:**  $\lim_{x \rightarrow \infty} (1 + \frac{a}{x})^x = e^a$

**NOTE:**

**field:** <sub>101</sub>

**field:**  $\lim_{x \rightarrow \infty} (1 + \frac{a}{x})^x = e^a$

**field:**  $\lim_{x \rightarrow \infty} (1 + \frac{a}{x})^x =$

**NOTE:**

**field:**  $_{102}$

**field:**  $\frac{d}{dx} f(g(x)) =$

**field:**  $\frac{d}{dx} f(g(x)) = f'(g(x))g'(x)$  (Chain rule)

**NOTE:**

**field:**  $_{103}$

**field:**  $\frac{d}{dx} \int_a^x f(t) dt =$

**field:**  $\frac{d}{dx} \int_a^x f(t) dt = f(x)$  (fundamental theorem of calculus )

**NOTE:**

**field:**  $_{104}$

**field:**  $\int_a^b u dv =$   
**ex:**  $\int x e^{-x}$

**field:**  $\int_a^b u dv = uv|_a^b - \int_a^b v du$   
**ex:**  $u = x, dv = e^{-x}, du = dx, v = -e^{-x}$

$$\begin{aligned} \int x e^{-x} &= -x e^{-x} + \int e^{-x} \\ &= -x e^{-x} - e^{-x} + c \end{aligned}$$

**NOTE:**

**field:** <sub>105</sub>

**field:**  $\sum_{k=0}^{\infty} \frac{x^k}{k!} =$

**field:**  $\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$

**NOTE:**

**field:** <sub>106</sub>

**field:**  $e^x =$

**field:**  $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

**NOTE:**

**field:** <sub>107</sub>

**field:**  $\sum_{k=0}^{\infty} x^k =$

**field:**  $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$  for  $|x| < 1$

**NOTE:**

**field:** <sub>108</sub>

**field:**  $\sum_{k=0}^n x^k =$

**field:**  $\sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x}$  for  $x \neq 1$

**NOTE:**

**field:** <sub>109</sub>

**field:**  $\lim_{x \rightarrow -\infty} e^{-x} =$



**field:**  $\lim_{x \rightarrow -\infty} e^{-x} = \infty$

**NOTE:**

**field:** <sub>110</sub>

**field:**  $\lim_{x \rightarrow \infty} e^{-x} =$

**field:**  $\lim_{x \rightarrow -\infty} e^{-x} = 0$

**NOTE:**

**field:** <sub>111</sub>

**field:**

$$(fg)' =$$

**field:**

$$(fg)' = f'g + g'f$$

(product rule )

**NOTE:**

**field:** <sub>112</sub>

**field:**  $\frac{d}{dx} x^n =$

**field:**  $\frac{d}{dx} x^n = nx^{n-1}$

**NOTE:**

**field:** <sub>113</sub>

**field:**  $\frac{d}{dx} a^x =$

**field:**  $\frac{d}{dx}a^x = a^x \ln(a)$

**NOTE:**

**field:**  $_{114}$

**field:**  $\frac{d}{dx} \ln(x) =$

**field:**  $\frac{d}{dx} \ln(x) = \frac{1}{x}$

**NOTE:**

**field:**  $_{115}$

**field:**  $\frac{d}{dx}(f(x))^n =$

**field:**  $\frac{d}{dx}(f(x))^n = n(f(x))^{n-1}f'(x)$

**NOTE:**

**field:**  $_{116}$

**field:**  $\frac{d}{dx} \ln(f(x)) =$

**field:**  $\frac{d}{dx} \ln(f(x)) = \frac{f'(x)}{f(x)}$

**NOTE:**

**field:**  $_{117}$

**field:**  $\frac{d}{dx} e^{f(x)} =$

**field:**  $\frac{d}{dx} e^{f(x)} = f'(x)e^{f(x)}$

**NOTE:**

**field:** <sub>118</sub>

**field:**  $\int x^n =$

**field:**  $\int x^n = \frac{1}{n+1}x^{n+1}$

**NOTE:**

**field:** <sub>119</sub>

**field:**  $\int \frac{1}{x} =$

**field:**  $\int \frac{1}{x} = \ln(|x|)$

**NOTE:**

**field:** <sub>120</sub>

**field:**  $\int \frac{1}{ax+b} =$

**field:**  $\int \frac{1}{ax+b} = \frac{1}{a}\ln(|ax+b|)$

**NOTE:**

**field:** <sub>121</sub>

**field:**  $\int e^{cx} =$

**field:**  $\int e^{cx} = \frac{1}{c}e^{cx}$

**NOTE:**

**field:** <sub>122</sub>

**field:**  $\int xe^{-cx^2} =$

**field:**  $\int x e^{-cx^2} = -\frac{1}{2c} e^{-cx^2}$

**NOTE:**

**field:** <sub>123</sub>

**field:** U substitution:  
example;  $\int_1^2 5x^2 \cos(x^3)$

**field:**  $\int_a^b f(g(x))g'(x) = \int_{g(a)}^g(b)f(u)du$   
Where  $u = g(x), du = g'dx$   
Ex:  $u = x^3, du = 3x^2, x^2du = 1/3du$   $\int_1^2 5x^2 \cos(x^3) = \int_1^8 5/3 \cos(u)du$

**NOTE:**

**field:** <sub>124</sub>

**field:**  $\Gamma(a) =$

**field:**  $\int_0^\infty t^{a-1} e^{-t} dt$

**NOTE:**

**field:** <sub>125</sub>

**field:**  $\int_0^\infty t^{a-1} e^{-t} dt$

**field:**  $= \Gamma(a)$

**NOTE:**

**field:** <sub>126</sub>

**field:**  $\Gamma(a + 1) =$

**field:**  $\Gamma(a + 1) = a\Gamma(a)$

**NOTE:**

**field:** <sub>127</sub>

**field:**  $\Gamma(n) =$

**field:**  $\Gamma(n) = (n - 1)!$  (for  $n$  an integer)

**NOTE:**

**field:** <sub>128</sub>

**field:**  $\Gamma(1/2) =$

**field:**  $\Gamma(1/2) = \sqrt{\pi}$

**NOTE:**

**field:** <sub>129</sub>

**field:**  $\Gamma(1) =$

**field:**  $\Gamma(1) = 1$

**tags:** Theory1

**NOTE:**

**field:** <sub>130</sub>

<b>field:</b>		
	number of trials Draw till nth success	replace   no replacement

field:		replace	no replacement
	number of trials	Binom	Hypergeometric
	Draw till nth success	Nbinom	Negative hypergeometric

**NOTE:**

**field:** <sup>131</sup>

**field:** Plug uniform into inverse CDF

**field:** Get cdf

**NOTE:**

**field:** <sup>132</sup>

**field:** Sample Space

**field:** The set,  $S$ , of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

**NOTE:**

**field:** <sup>133</sup>

**field:** Event

**field:** An *event* is any collection of possible outcomes of an experiment, that is, any subset of  $S$  (including  $S$  itself).

**NOTE:**

**field:** <sup>134</sup>

**field:** Union

**field:**  $A \cup B = \{x : x \in A \text{ or } x \in B\}$

**NOTE:**

**field:** <sub>135</sub>

**field:** Intersection

**field:**  $A \cap B = \{x : x \in A \text{ and } x \in B\}$

**NOTE:**

**field:** <sub>136</sub>

**field:** Complementation

**field:**  $A^c = \{x : x \notin A\}$

**NOTE:**

**field:** <sub>137</sub>

**field:** Commutativity

$$A \cup B =$$

$$A \cap B =$$

**field:** Commutativity

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

**NOTE:**

**field:** <sub>138</sub>

**field:** Associativity

$$A \cup (B \cup C) =$$

$$A \cap (B \cap C) =$$

**field:** Associativity

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

**NOTE:**

**field:** <sub>139</sub>

**field:** Distributive Laws

$$A \cap (B \cup C) =$$

$$A \cup (B \cap C) =$$

**field:** Distributive Laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

**NOTE:**

**field:** <sub>140</sub>



**field:** DeMorgan's Laws

$$(A \cup B)^c =$$

$$(A \cap B)^c =$$

**field:** DeMorgan's Laws

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

**NOTE:**

**field:** <sub>141</sub>

**field:** Disjoint

**field:** Disjoint: Two events  $A$  and  $B$  are disjoint ( or mutually exclusive) if  $A \cap B = \emptyset$

**NOTE:**

**field:** <sub>142</sub>

**field:**

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) =$$

**field:**

$$P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1 \cdots A_{n-1})$$

**NOTE:**

**field:** <sup>143</sup>

**field:**

$$P(A, B, C) =$$

**field:**

$$P(A, B, C) = P(A)P(B|A)P(C|A, B)$$

**NOTE:**

**field:** <sup>144</sup>

**field:**

$$P(A \cup B \cup C) =$$

**field:**

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

**NOTE:**

**field:** <sup>145</sup>

**field:** Pairwise disjoint

**field:** Two Events  $A_1, A_2$  are pairwise disjoint ( or mutually exclusive) if  $A_i \cap A_j = \emptyset$  for all  $i \neq j$

**NOTE:**

**field:** <sup>146</sup>

**field:** Partition

**field:** If  $A_1, A_2, \dots$  are pairwise disjoint and  $\cup_{i=1}^{\infty} A_i = S$ , then the collection  $A_1, A_2, \dots$  forms a partition of  $S$ .

**NOTE:**

**field:** <sup>147</sup>

**field:** Sigma Algebra

**field:** A collection of subsets of  $S$  is called a sigma algebra (or Borel field), denoted by  $\mathcal{B}$ , if it satisfies the following three properties:

1.  $\emptyset \in \mathcal{B}$  (the empty set is an element of  $\mathcal{B}$ )
2. If  $A \in \mathcal{B}$ , then  $A^c \in \mathcal{B}$  ( $\mathcal{B}$  is closed under complementation)
3. If  $A_1, A_2, \dots \in \mathcal{B}$ , then  $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$  ( $\mathcal{B}$  is closed under countable unions)

**NOTE:**

**field:** <sup>148</sup>

**field:** Probability Function / Kolmogorov Axioms

**field:** Given a sample space  $S$  and an associated sigma algebra  $\mathcal{B}$ , a probability function is a function  $P$  with domain  $\mathcal{B}$  that satisfies:

1.  $P(A) \geq 0$  for all  $A \in \mathcal{B}$
2.  $P(S) = 1$
3. If  $A_1, A_2, \dots \in \mathcal{B}$  are pairwise disjoint, then  $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$   
(Axiom of Countable Additivity)

**NOTE:**

**field:** <sub>149</sub>

**field:** If  $A \in \mathcal{B}$  and  $B \in \mathcal{B}$  are disjoint, then

$$P(A \cup B) = P(A) + P(B)$$

Axiom of Finite Additivity

**field:** If  $A \in \mathcal{B}$  and  $B \in \mathcal{B}$  are disjoint, then

$$P(A \cup B) = P(A) + P(B)$$

**NOTE:**

**field:** <sub>150</sub>

**field:** Properties of probability functions

1.  $P(\emptyset) =$
2.  $P(A)$
3.  $P(A^c) =$

**field:** Properties of probability functions

1.  $P(\emptyset) = 0$
2.  $P(A) \leq 1$
3.  $P(A^c) = 1 - P(A)$

**NOTE:**

**field:** <sub>151</sub>

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then

$$P(B \cap A^c) =$$

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then

$$P(B \cap A^c) = P(B) - P(A \cap B)$$

**NOTE:**

**field:** <sub>152</sub>

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then

$$P(A \cup B) =$$

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

**NOTE:**

**field:** <sub>153</sub>

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then if  $A \subset B$  then

**field:** If  $P$  is a probability function and  $A$  and  $B$  are any sets in  $\mathcal{B}$ , then if  $A \subset B$  then  $P(A) \leq P(B)$

**NOTE:**

**field:** 154

**field:** Bonferroni's Inequality

$$P(A \cap B)$$

**field:** Bonferroni's Inequality:

$$P(A \cap B) \geq P(A) + P(B) - 1$$

**NOTE:**

**field:** 155

**field:** If  $P$  is a probability function, then for any partition  $C_1, C_2, \dots$   $P(A) =$

**field:** If  $P$  is a probability function, then for any partition  $C_1, C_2, \dots$   $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$

**NOTE:**

**field:** 156

**field:** Boole's Inequality

$$P(\cup_{i=1}^{\infty} A_i)$$

**field:** If  $P$  is a probability function,

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i) \text{ for any sets } A_1, A_2, \dots$$

**NOTE:**

**field:** 157

**field:** Fundamental Theorem of Counting

**field:** If a job consists of  $k$  separate tasks, the  $i$ th of which can be done in  $n_i$  ways,  $i = 1, \dots, k$ , then the entire job can be done in  $n_1 \times n_2 \times \dots \times n_k$  ways.

**NOTE:**

**field:** 158

**field:** Ordered without replacement: number of arrangements of size  $r$  from  $n$  objects

**field:**

$$\frac{n!}{(n-r)!}$$

eg lottery with  $n = 44$  choices for  $r = 6$  values, cant use same number twice, order matters

**NOTE:**

**field:** 159

**field:** Unordered without replacement: number of arrangements of size  $r$  from  $n$  objects

**field:**

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

eg lottery with  $n = 44$  choices for  $r = 6$  values, cant use same number twice, order does not matter (Use ordered without replacement and divide by redundant orderings )

**NOTE:**

**field:** <sub>160</sub>

**field:** Ordered with replacement: number of arrangements of size  $r$  from  $n$  objects

**field:** Ordered with replacement: number of arrangements of size  $r$  from  $n$  objects

$$n^r$$

eg lottery with  $n = 44$  choices for  $r = 6$  values, can use same number twice, order matters

**NOTE:**

**field:** <sub>161</sub>

**field:** Unordered with replacement: number of arrangements of size  $r$  from  $n$  objects

**field:** Unordered with replacement: number of arrangements of size  $r$  from  $n$  objects

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

eg lottery with  $n = 44$  choices for  $r = 6$  values, can use same number twice, order does not matters

**NOTE:**



**field:** <sub>162</sub>

**field:** Number of arrangements of size  $r$  from  $n$  objects

	Without Replacement	With replacement
Ordered		
Unordered		

**field:** Number of arrangements of size  $r$  from  $n$  objects

	Without Replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	$n^r$
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

**NOTE:**

**field:** <sub>163</sub>

**field:** Binomial Coefficient  $\binom{n}{r}$

**field:** Binomial Coefficient

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

**NOTE:**

**field:** <sub>164</sub>

**field:**

$$P(A|B) =$$

**field:**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

**NOTE:**

**field:** <sub>165</sub>

**field:** Statistically independent  $P(A \cap B) =$

**field:** Statistically independent  $P(A \cap B) = P(A)P(B)$

**NOTE:**

**field:** <sub>166</sub>

**field:** If  $A$  and  $B$  are independent events, what else is independent?

**field:**

- $A$  and  $B^c$
- $A^c$  and  $B$
- $A^c$  and  $B^c$

**NOTE:**

**field:** <sub>167</sub>

**field:** Mutually independent

**field:** A collection of events  $A_1, \dots, A_n$  are mutually independent for any subcollection  $A_{i_1}, \dots, A_{i_k}$ , we have

$$P((\cap_{j=1}^k A_{ij})) = \prod_{j=1}^k P(A_{ij})$$

**NOTE:**

**field:** 168

**field:** Random variable

**field:** A random variable is a function from a sample space  $S$  into the real numbers

**NOTE:**

**field:** 169

**field:** Definition of a pdf

**field:** A function  $f_X(x)$  is a pdf (or pmf) of a random variable  $X$  if and only if

1.  $f_X(x) \geq 0$  for all  $x$
2.  $\sum_x f_X(x) = 1$  or  $\int_{-\infty}^{\infty} f_X(x)dx = 1$

**NOTE:**

**field:** 170

**field:** (Theorem) Let  $X$  have cdf  $F_X(x)$ , let  $Y = g(X)$

1. If  $g$  is an increasing function on  $X$ ,  $F_Y(y) = F_X(g^{-1}(y))$  for  $y \in Y$
2. If  $g$  is a decreasing function on  $X$  and  $X$  is a continuous random variable,  $F_Y(y) = 1 - F_X(g^{-1}(y))$  for  $y \in Y$

**field:** (Theorem) Let  $X$  have cdf  $F_X(x)$ , let  $Y = g(X)$

1. If  $g$  is an increasing function on  $X$ ,  $F_Y(y) = F_X(g^{-1}(y))$  for  $y \in Y$
2. If  $g$  is a decreasing function on  $X$  and  $X$  is a continuous random variable,  $F_Y(y) = 1 - F_X(g^{-1}(y))$  for  $y \in Y$

**NOTE:**

**field:** 171

**field:** Method of pdf

**field:** Conditions:

1.  $g$  is a monotone function
2.  $f_X(x)$  is continuous on  $X$
3.  $g^{-1}(y)$  has a continuous derivative

Let  $X$  have pdf  $f_x(x)$  and let  $Y = g(X)$

$$f_Y(y) = f_x(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

**NOTE:**

**field:** 172

**field:** (Theorem) Let  $X$  have cdf  $F_X(x)$ , let  $Y = g(X)$

- If  $g$  is an increasing function,  $F_Y(y) =$
- If  $g$  is a decreasing function, and  $X$  is a continuous random variable,  $F_Y(y) =$

**field:** (Theorem) Let  $X$  have cdf  $F_X(x)$ , let  $Y = g(X)$

- If  $g$  is an increasing function,  $F_Y(y) = F_X(g^{-1}(y))$
- If  $g$  is a decreasing function, and  $X$  is a continuous random variable,  $F_Y(y) = 1 - F_X(g^{-1}(y))$

**NOTE:**

**field:** <sub>173</sub>

**field:** eg:  $X \sim Unif(0, 1)$ ,  $Y = -\log(X)$   $F_Y(y) =$

**field:**  $F_Y(y) = 1 - F_x(g^{-1}(y)) = 1 - F_X(e^{-y}) = 1 - e^{-y}$

**NOTE:**

**field:** <sub>173x</sub>

**field:**  $X$  is a continuous random variable. For  $y > 0$ ,  $Y = X^2$   $F_Y(y) =$

**field:**

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

**NOTE:**

**field:** <sub>174</sub>

**field:** Pdf of  $F_X(g(X))$ , where  $Y = g(X)$

**field:** Chain rule:  $f_Y(y) = g'(y)f(g(y))$

**NOTE:**

**field:** <sub>175</sub>

**field:** Method of pdf if  $g$  is not monotone all entire domain

**field:**  $f_Y = \sum f_x(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|$   $y \in Y$ , 0 otherwise  
 eg:  $Y = X^2$ ,

**NOTE:**

**field:** <sub>176</sub>

**field:**  $P(Y \leq y)$  when  $Y = F_X(x)$

**field:**

$$\begin{aligned} P(Y \leq y) &= P(X \leq F_x^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

$Y$  is uniformly distributed

**NOTE:**

**field:** <sub>177</sub>

**field:**  $M_x(t) = (\text{discrete})$

**field:**  $M_x(t) = E(e^{tX}) = \sum_x e^{tX} P(X)$  (discrete)

**NOTE:**

**field:** <sub>178</sub>

**field:**  $M_x(t) = (\text{continuous})$

**field:**  $M_x(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f_x(x) dx$  (continuous)

**NOTE:**

**field:** <sub>179</sub>

**field:**  $E(X^n) =$

**field:**  $E(X^n) = M_x^n(0) = \frac{d^n}{dt^n} M_x(t)|_{t=0}$

**NOTE:**

**field:** <sub>180</sub>

**field:**  $M(aX + b)(t) =$

**field:**  $M(aX + b)(t) = e^{bt} M_x(at)$

**NOTE:**

**field:** <sub>181</sub>

**field:** If  $E(X^n)$  exists then...

**field:** If  $E(X^n)$  exists then  $E(X^m)$  exists for  $m \leq n$

**NOTE:**

**field:** <sub>182</sub>

**field:** If  $X_i$  are independent and  $Y = a_1X_1 + \dots + a_nX_n + b$ , then  $M_Y(t) =$

**field:** If  $X_i$  are independent and  $Y = a_1X_1 + \dots + a_nX_n$ , then  $M_Y(t) = e^{bt} \prod_{i=1}^n M_{X_i}(a_it)$

**NOTE:**

**field:** <sub>183</sub>

**field:** Example of using MGF for finding expected value: MGF gamma:  $(\frac{1}{1-\beta t})^\alpha$ :  $E(X) =$

**field:**  $E(X) = \frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}|_{t=0} = \alpha\beta$

**NOTE:**

**field:** 184

**field:** Using MGF to relate distributions: MGF exp =  $(1 - \beta t)^{-1}$

**field:**  $Y = \sum X_i$  is gamma as MGF gamma is  $(1 - \beta t)^{-\alpha}$

**NOTE:**

**field:** 185

**field:** First step in transforming a RV

**field:** Determine support

**NOTE:**

**field:** 186

**field:**  $n$ th Moment of  $X$

**field:**  $E(X^n)$

**NOTE:**

**field:** 187

**field:**  $n$ th central moment of  $X$

**field:**  $E(X - \mu)^n$

**NOTE:**



**field:** <sub>188</sub>

**field:**  $(a + b)^n =$

**field:**  $(a + b)^n = \sum_{x=0}^n \binom{n}{x} a^x b^{n-x}$

**NOTE:**

**field:** <sub>189</sub>

**field:**  $\sum_{x=0}^n \binom{n}{x} a^x b^{n-x} =$

**field:**  $(a + b)^n$

**NOTE:**

**field:** <sub>190</sub>

**field:**  $N$  balls  $r$  red  $N - r$  green. Select  $n$  balls. Probability that  $y$  are red?

**field:** Hypergeometric distribution( $N, r, n$ )

**NOTE:**

**field:** <sub>191</sub>

**field:** Hypergeometric distribution description ( $N, r, n$ )

**field:**  $N$  is total balls,  $r$  is number red balls,  $n$  is number balls selected.

**NOTE:**

**field:** <sub>192</sub>

**field:** Negative binomial description

**field:** Number of Bernoulli trials required to get a fixed number of successes.  
 $r$  being the  $r$ th success

**NOTE:**

**field:** <sub>193</sub>

**field:** Geometric description

**field:** Modeling waiting time.  $X$  is the trial at which the first success occurs.

**NOTE:**

**field:** <sub>194</sub>

**field:** Location-scale family for  $f(x)$

**field:**  $1/\sigma f((x - \mu)/\sigma)$

**NOTE:**

**field:** <sub>195</sub>

**field:** Given  $X$  give the mean and variance for the location-scale random  
 $Y = 1/\sigma f((y - \mu)/\sigma)$  variable

**field:**  $E(Y) = \sigma E(X) + \mu, V(Y) = \sigma^2 V(X)$

**NOTE:**

**field:** <sub>196</sub>

**field:**  $X \sim Pois(\lambda) \ P(X = x + 1) =$

**field:**  $X \sim Pois(\lambda) \ P(X = x + 1) = \frac{\lambda}{x+1} P(X = x)$

**NOTE:**

**field:** <sub>197</sub>

**field:**  $f(y|x) =$  (definition of conditional)

**field:**  $f(y|x) = \frac{f(x,y)}{f_x(x)}$

**NOTE:**

**field:** <sub>198</sub>

**field:**  $E(g(Y)|x) =$

**field:**  $E(g(Y)|x) = \int_{-\infty}^{\infty} g(y)f(y|x)dy$

**NOTE:**

**field:** <sub>199</sub>

**field:** Example of calculating conditional pdfs  $f(x, y) = e^{-y}, 0 < x < y < \infty$ .  $f(y|x) =$

**field:**

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy \\ = e^{-x}$$

$$f(y|x) = \frac{f(x, y)}{f_x(x)} \\ = \frac{e^{-y}}{e^{-x}} \text{ if } y > x \\ = \frac{0}{e^{-x}} \text{ if } y \leq x$$

**NOTE:**

**field:** 200

**field:** Let  $(X, Y)$  be given as  $f(x, y)$ . Then  $X$  and  $Y$  are independent if

**field:** Let  $(X, Y)$  be given as  $f(x, y)$ . Then  $X$  and  $Y$  are independent if there exist functions  $g(x), h(y)$  such that  $f(x, y) = g(x)h(y)$  (factorization - don't need to compute marginals )

**NOTE:**

**field:** 201

**field:** Let  $X, Y$  be independent. Then  $E(g(X)h(Y)) =$

**field:**  $E(g(X)h(Y)) = (E(g(X)))(E(h(Y)))$   
example:  $E(X^2Y) = E(X^2)E(Y)$

**NOTE:**

**field:** 202

**field:**  $X, Y$  independent

$$Z = X + Y$$

$$M_Z(t) =$$

$$\text{field: } M_Z(t) = M_X(t)M_Y(t)$$

**NOTE:**

**field:** <sub>203</sub>

**field:** Method of pdf bivariate

$$\text{field: } f_{u,v}(u, v) = f_{x,y}(h_1(u, v), h_2(u, v))|J|$$

$$\text{Where } |J| = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial y}{\partial u} \frac{\partial x}{\partial v}$$

$$\text{and } u = g_1(x, y), v = g_2(x, y) \text{ and } x = h_1(x, y), y = h_2(x, y)$$

**NOTE:**

**field:** <sub>204</sub>

**field:**  $X, Y$  independent,  $g(X)$  a function only of  $X$  and  $h(Y)$  a function only of  $Y$ . Then

**field:**  $g(X)$  and  $g(Y)$  are independent.

**NOTE:**

**field:** <sub>205</sub>

**field:** Correlation

$$\text{field: } \rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

**NOTE:**

**field:** 206

**field:**  $m$  independent trials, each trial resulting in one of  $n$  outcomes, with probabilities  $p_1, \dots, p_n$ .  $X_i$  is the count of the number of times the  $i$ th outcome occurred in the  $m$  trials.

**field:** Multinomial distribution  $f(x_1, \dots, x_n) = \frac{m!}{x_1! \dots x_n!} p_1^{x_1} \dots p_n^{x_n}$

**NOTE:**

**field:** 207

**field:**  $|E(XY)|$

**field:**  $|E(XY)| \leq E(|XY|) \leq (E(|X|^2))^{1/2} (E(|Y|^2))^{1/2}$  (Cauchy-Schwartz)

**NOTE:**

**field:** 208

**field:**  $E(g(X)) \geq$  where  $g$  is a convex function

**field:**  $E(g(X)) \geq g(E(X))$  where  $g$  is a convex function (Jensen's inequality)

**NOTE:**

**field:** 209

**field:** Ranking of types of means

**field:**  $\mu_{\text{harmonic}} \leq \mu_{\text{geometric}} \leq \mu_{\text{arithmetic}}$  By Jensen's inequality (using logs)

**NOTE:**

**field:** <sub>210</sub>

**field:** Linear transformations of multivariate normal  $X \sim N(\vec{\mu}, \Sigma)$   
 $A\vec{X} + \vec{b}$

**field:**  $A\vec{X} + \vec{b} \sim N(A\vec{\mu} + \vec{v}, A\Sigma A^t)$

**NOTE:**

**field:** <sub>211</sub>

**field:**  $X \sim N(\vec{\mu}, \Sigma)$

$$\vec{X}_a | \vec{X}_b \sim$$

**field:**  $\vec{X}_a | \vec{X}_b \sim N(\vec{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\vec{x}_b - \vec{\mu}_b), \Sigma_{ba} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})$

$$\text{ex: } (X_1, X_2, X_3), \vec{\mu} = (1, 2, 3)^t, \Sigma = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix} \quad X_1, X_3 | X_2 = 1$$

$$a = \{1, 3\}, b = \{2\}$$

$$\mu_a = (1, 3)^t, \mu_b = 1$$

$$\Sigma_{aa} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}, \Sigma_{ab} = (1, 1)^t$$

**NOTE:**

**field:** <sub>212</sub>

**field:**  $(X, Y)$  multinomial  
 $aX + bY \sim$

**field:**  $aX + bY \sim N(a\mu_x + b\mu_y, a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\rho\sigma_x\sigma_y)$

**NOTE:**

**field:** <sub>213</sub>

**field:**  $(X, Y)$  multinomial  
 $Y|X \sim$

**field:**  $Y|X \sim N(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_Y^2(1 - \rho^2))$

**NOTE:**

**field:** <sub>214</sub>

**field:** CDF for Max order statistic

**field:**  $(F(x))^n$

**NOTE:**

**field:** <sub>215</sub>

**field:** PDF for Max order statistic

**field:**  $n(F(x))^{n-1}f(x)$

**NOTE:**

**field:** <sub>216</sub>

**field:** CDF for Min order statistic

**field:**  $1 - (1 - F(x))^n$

**NOTE:**

**field:** <sub>217</sub>



**field:** PDF for Min order statistic

**field:**  $n(1 - F(x))^{n-1}f(x)$

**NOTE:**

**field:** <sub>218</sub>

**field:** CDF for  $k$ th order statistic

**field:**  $F_{(k)}(x) = \sum_{j=k}^n \binom{n}{j} (F(x))^j (1 - F(x))^{n-j}$

**NOTE:**

**field:** <sub>219</sub>

**field:** PDF for  $k$ th order statistic

**field:**  $f_{(k)}(x) = k \binom{n}{k} f(x) F(x)^{k-1} (1 - F(x))^{n-k}$

**NOTE:**

**field:** <sub>1</sub>

**field:** CDF of Geometric ( $p$ )

**field:**  $1 - (1 - p)^x$

**NOTE:**

**field:** <sub>2</sub>

**field:** CDF of Exponential( $\beta$ )

**field:**  $1 - e^{-\frac{x}{\beta}}$

**NOTE:**

**field:** <sub>3</sub>

**field:**

- $P(\emptyset) =$
- $B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$
- $P(A^c) =$
- $P(B) =$
- $P(\Omega) = \quad P(\emptyset) =$
- $(\bigcup_n A_n) = \quad (\bigcap_n A_n) = \quad \text{DEMORGAN}$

**field:**

- $P(\emptyset) = 0$
- $B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$
- $P(A^c) = 1 - P(A)$
- $P(B) = P(A \cap B) + P(A^c \cap B)$
- $P(\Omega) = 1 \quad P(\emptyset) = 0$
- $(\bigcup_n A_n) = \bigcap_n A_n \quad (\bigcap_n A_n) = \bigcup_n A_n \quad \text{DEMORGAN}$

**NOTE:**

**field:** <sub>4</sub>

**field:** Probability Set intersection

- $P(\bigcup_n A_n) = 1 - P(\bigcap_n A_n^c)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) \implies P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B) =$
- $P(A \cap B^c) =$

**field:** Probability Set intersection

- $P(\bigcup_n A_n) = 1 - P(\bigcap_n A_n^c)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$   
 $\implies P(A \cup B) \leq P(A) + P(B)$
- $P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$
- $P(A \cap B^c) = P(A) - P(A \cap B)$

**NOTE:**

**field:** 5

**field:**  $P(A \cap B) = P(A)P(B)$  when  $A$  and  $B$  independent

**field:**  $P(A \cap B) = P(A)P(B)$  when  $A$  and  $B$  independent

**NOTE:**

**field:** 6

**field:**

$$P(A|B) =$$

**field:**

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

**NOTE:**

**field:** 7

**field:** Law of total probability

**field:** Law of total probability

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i) \quad \Omega = \cup_{i=1}^n A_i$$

$$P(B) = P(A \cup B) + P(A^c \cup B)$$

**NOTE:**

**field:** 8

**field:** Bayes Theorem

**field:** Bayes Theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)} \quad \Omega = \cup_{i=1}^n A_i$$

**NOTE:**

**field:** 9

**field:** CDF Laws

**field:** CDF Laws

1. Nondecreasing:  $x_1 < x_2 \implies F(x_1) \leq F(x_2)$
2. Limits:  $\lim_{x \rightarrow -\infty} F(x) = 0$  and  $\lim_{x \rightarrow \infty} F(x) = 1$
3. Right-Continuous  $\lim_{y \rightarrow x^+} F(y) = F(x)$

**NOTE:**

**field:** 10

**field:**

$$f_{y|x}(y|x) =$$

**field:**

$$f_{y|x}(y|x) = \frac{f(x, y)}{f_x(x)}$$

**NOTE:**

**field:** <sub>11</sub>

**field:**  $X, Y$  independent

- $P(X \leq x, Y \leq y) =$
- $f_{x,y}(x, y) =$

**field:**  $X, Y$  independent

- $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$
- $f_{x,y}(x, y) = f_x(x)f_y(y)$

**NOTE:**

**field:** <sub>12</sub>

**field:** Transformations  $Z = \phi(X)$

- Discrete:  $f_Z(z) =$
- Continuous:  $F_Z(z) =$
- Cont,  $\phi$  strictly monotone:  $f_z(z)$

**field:** Transformations  $Z = \phi(X)$

- Discrete:

$$f_Z(z) = P(\phi(X) = z) = P(X \in \phi^{-1}(z)) = \sum_{x \in \phi^{-1}(z)} f_x(x)$$

- Continuous (Method of CDF):

$$F_Z(z) = P(\phi(X) \leq z) = \int_{x: \phi(x) \leq z} f(x) dx$$

- Cont,  $\phi$  strictly monotone: (Method of PDF)  $f_z(z) = f_x(\phi^{-1}(z)) \left| \frac{d}{dz} \phi^{-1}(z) \right|$

**NOTE:**

**field:** <sub>13</sub>

**field:** Rule of the Lazy Statistician:  $E[g(x)] =$

**field:** Rule of the Lazy Statistician:  $E[g(x)] = \int g(x) f_x(x) dx$

**NOTE:**

**field:** <sub>14</sub>

**field:** Expectation rules

- $E(c) =$
- $E(cX) =$
- $E(X + Y) =$
- $E(\phi(X)) =$

**field:** Expectation rules

- $E(c) = c$
- $E(cX) = cE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $E(\phi(X)) \neq \phi(E(X))$

**NOTE:**

**field:** <sub>15</sub>

**field:** Conditional expectation

- $E(Y|X = x) =$
- $E(X) =$
- $E(Y + Z|X) =$
- $E(Y|X) = c \implies$

**field:** Conditional expectation

- $E(Y|X = x) = \int yf(y|x)dy$
- $E(X) = E(E(X|Y))$
- $E(Y + Z|X) = E(Y|X) + E(Z|X)$
- $E(Y|X) = c \implies Cov(X, Y) = 0$

**NOTE:**

**field:** <sub>16</sub>

**field:** Variance

- $V(X) = \sigma_x^2 =$
- $V(X + Y) =$
- $V\left[\sum_{i=1}^n X_i\right] =$

**field:** Variance

- $V(X) = \sigma_x^2 = E[(X - E(X))^2] = E(X^2) - E(X)^2$
- $V(X + Y) = V(X) + V(Y) + Cov(X, Y)$
- $V\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n V(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$

**NOTE:**

**field:** <sub>17</sub>

**field:** Covariance

- $Cov(X, Y) =$
- $Cov(X, c) =$
- $Cov(Y, X) =$
- $Cov(aX, bY) =$
- $Cov(X + a, Y + b) =$
- $Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) =$



**field:** Covariance

- $Cov(X, Y) = E[(X - E(X))(Y - E(Y)))] = E(XY) - E(X)E(Y)$
- $Cov(X, c) = 0$
- $Cov(Y, X) = Cov(X, Y)$
- $Cov(aX, bY) = abCov(X, Y)$
- $Cov(X + a, Y + b) = Cov(X, Y)$
- $Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j)$

**NOTE:**

**field:** <sub>18</sub>

**field:** Correlation:  $\rho(X, Y)$

**field:** Correlation:  $\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$

**NOTE:**

**field:** <sub>19</sub>

**field:** Conditional Variance

- $V(Y|X) =$
- $V(Y) =$

**field:** Conditional Variance

- $V(Y|X) = E[(Y - E(Y|X))^2|X] = E(Y^2|X) - E(Y|X)^2$
- $V(Y) = E(V(Y|X)) + V(E(Y|X))$

**NOTE:**

**field:** <sub>20</sub>

**field:** Law of total probability  $k = 2$  (using conditional probability)

**field:**  $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

**NOTE:**

**field:** <sub>21</sub>

**field:** Bayes formula in terms of law of total probability,

**field:**  $P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$

**NOTE:**

**field:** <sub>22</sub>

**field:**  $P(A \text{ and } B) = P(A, B) =$

**field:**  $P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$

**NOTE:**

**field:** <sub>23</sub>

**field:** Events  $A$  and  $B$  are independent if

**field:**  $P(A|B) = P(A)$  equivalently  $P(A \text{ and } B) = P(A)P(B)$

**NOTE:**

**field:** <sub>24</sub>

**field:** Poisson setting

**field:** The Poisson setting arises in the context of discrete counts of events that occur over space or time with the small probability and where successive events are independent

Eg: 2 on average calls a minute,  $X$  is number of calls a minute,  $X \sim Pois$

**NOTE:**

**field:** <sub>25</sub>

**field:** Poisson approximation of binomial distribution

**field:** Suppose  $X \sim Binom(n, p)$ ,  $Y \sim Pois(\lambda)$ . If  $n \rightarrow \infty$ , and  $p \rightarrow 0$ , in such a way that  $np \rightarrow \lambda > 0$ , then for all  $k$ ,  $P(X = k) \rightarrow P(Y = k)$ . The Poisson distribution with parameter  $\lambda = np$  serves as a good approximation for the binomial distribution when  $n$  is large and  $p$  is small.

**NOTE:**

**field:** <sub>26</sub>

**field:**  $E(f(X, Y))$  when  $X, Y$  are discrete

**field:**  $E(f(X, Y)) = \sum_x \sum_y f(x, y)P(X = x, Y = y)$

**NOTE:**

**field:** <sub>27</sub>

**field:** If  $X, Y$  are independent, then  $f(X), g(Y)$

**field:** are also independent

**NOTE:**

**field:** <sub>28</sub>

**field:** If  $X, Y$  independent,  $E(XY) = E(f(X)g(Y)) =$

**field:** If  $X, Y$  independent,  $E(XY) = E(X)E(Y), E(f(X)g(Y)) = E(f(X))E(g(Y))$

**NOTE:**

**field:** <sub>29</sub>

**field:** Sum of independent discrete random variables  $X, Y$ :  $P(X + Y = k)$

**field:**  $P(X + Y = k) = \sum_i P(X = i)P(Y = k - i)$

**NOTE:**

**field:** <sub>30</sub>

**field:**  $V(X) = 0$

**field:** If and only if  $X$  is a constant

**NOTE:**

**field:** <sub>31</sub>

**field:**  $E(I_A) =, V(I_A)$  Where  $I_A$  is an indicator function

**field:**  $E(I_A) = P(A), V(I_A) = P(A)P(A^c)$

**NOTE:**

**field:** <sub>32</sub>

**field:** For discrete jointly distributed random variables,

$$P(X = y|X = x) =$$

**field:** For discrete jointly distributed random variables,

$$P(X = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

**NOTE:**

**field:** <sub>33</sub>

**field:** For discrete random variables  $E(Y|X = x) =$

**field:** For discrete random variables  $E(Y|X = x) = \sum_y yP(Y = y|X = x)$

**NOTE:**

**field:** <sub>34</sub>

**field:** Problem solving strategy for expected value of counting

**field:** Use indicator functions for each trial , where  $X = \sum I$  and use linearity of expectation

**NOTE:**

**field:** <sub>35</sub>

**field:**  $P(X > s + t|X > t)$  for geometric, exponential

**field:**  $P(X > s + t|X > t) = P(X > s)$

**NOTE:**

**field:** <sub>36</sub>

**field:** Distribution for: A bag of  $N$  balls which contains  $r$  red balls and  $N - r$  blue balls,  $X$  is number of red balls in a sample of size  $n$  taken without replacement.

**field:** Hypergeometric.

**NOTE:**

**field:** <sub>37</sub>

**field:** Distribution for modeling arrival time

**field:** Exponential

**NOTE:**

**field:** <sub>38</sub>

**field:**  $E(g(X, Y)) = (\text{continuous})$

**field:**  $E(g(X, Y)) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} g(x, y) f(x, y) dx dy$

**NOTE:**

**field:** <sub>39</sub>

**field:**  $Cov(X, Y) = (\text{integration})$

**field:**  $Cov(X, Y) = \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{\infty} (x - E(X))(y - E(Y)) dx dy$

**NOTE:**

**field:** <sub>40</sub>

**field:** Problem solving strategies for functions of random variables

**field:**

- Methods of cdf:  $Y = g(X)$ , find cdf  $P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y))$
- For finding  $P(X < Y)$ , set up integrals that cover
- For finding probabilities of independent uniform random variables, use geometric (area) properties

**NOTE:**

**field:** <sub>41</sub>

**field:** Quantile

**field:** If  $X$  is a continuous random variable, then the  $p$ th quantile is the number  $q$  that satisfies  $P(X \leq q) = p/100$

**NOTE:**

**field:** <sub>42</sub>

**field:** Poisson process

**field:** Times between arrivals are modeled as iid exponential random variables with parameter  $\lambda = 1/\beta$ . Let  $N_t$  be the number of arrivals up to time  $t$ . Then  $N_t \sim Pois(\lambda t)$

**NOTE:**

**field:** <sub>43</sub>

**field:** Conditional density function  $f_{Y|X}(y|x) =$

**field:**  $f_{Y|X}(y|x) = \frac{f(x,y)}{f_x(x)}$

**NOTE:**

**field:** <sub>44</sub>

**field:** Continuous bayes formula

**field:**  $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_x(x)}{\int_{t=-\infty}^{\infty} f_{Y|X}(y|t)f_x(t)dt}$

**NOTE:**

**field:** <sub>45</sub>

**field:** Conditional expectation for continuous random variables  $E(Y|X = x)$

**field:**  $E(Y|X = x) = \int_y y f_{Y|X}(y|x) dy$

**NOTE:**

**field:** <sub>46</sub>

**field:** Law of total expectation

**field:**  $E(Y) = E(E(Y|X))$

**NOTE:**

**field:** <sub>47</sub>



**field:** Properties of conditional expectation

- $E(aY + bZ|X) =$
- $E(g(Y)|X = x) =$
- If  $X, Y$  independent,  $E(Y|X) =$
- If  $Y = g(X)$ , then  $E(Y|X) =$

**field:** Properties of conditional expectation

- $E(aY + bZ|X) = aE(Y|X) + bE(Z|X)$
- $E(g(Y)|X = x) = \int_y g(y)f_{Y|X}(y|x)dy$
- If  $X, Y$  independent,  $E(Y|X) = E(Y)$
- If  $Y = g(X)$ , then  $E(Y|X) = Y$

**NOTE:**

**field:** <sub>49</sub>

**field:** Conditional variance  $V(Y|X = x)$

**field:**

$$V(Y|X = x) = \sum_y (y - E(Y|X = x))^2 P(Y = y|X = x)$$

discrete

$$V(Y|X = x) = \int_y (y - E(Y|X = x))^2 f_{Y|X}(y|x)dy$$

continuous

**NOTE:**

**field:** <sub>50</sub>

**field:** Properties of conditional variance

- $V(Y|X = x) =$
- $V(aY + b|X = x) =$
- If  $Y, Z$  independent,  $V(Y + Z|X = x) =$

**field:** Properties of conditional variance

- $V(Y|X = x) = E(Y^2|X = x) - (E(Y|X = x))^2$
- $V(aY + b|X = x) = a^2V(Y|X = x)$
- If  $Y, Z$  independent,  $V(Y + Z|X = x) = V(Y|X = x) + V(Z|X = x)$

**NOTE:**

**field:** <sub>51</sub>

**field:**  $P(X \geq \epsilon)$

**field:**  $P(X \geq \epsilon) \leq E(X)/\epsilon$  (Markov's Inequality )

**NOTE:**

**field:** <sub>52</sub>

**field:**  $P(|X - \mu| \geq \epsilon)$

**field:**  $P(|X - \mu| \geq \epsilon) \leq \sigma^2/\epsilon^2$  (Chebyshev's inequality, if mean and variance finite )

**NOTE:**

**field:** <sub>53</sub>

**field:**  $P(\lim_{n \rightarrow \infty} S_n/n = \mu) =$

**field:**  $P(\lim_{n \rightarrow \infty} S_n/n = \mu) = 1$  (Strong law of large numbers )

**tags:** TheoryTwo t2

**NOTE:**

**field:** 220

**field:** Definition of Convergence

**field:** A sequence  $\{a_n\}_{n \geq 1}$  of real numbers is said to **converge** to a point  $a \in \mathbb{R}$  if for any  $\epsilon > 0$  there exists  $N \in \mathbb{N}$  such that for all  $m > N$  we have  $|a_m - a| < \epsilon$

**NOTE:**

**field:** 221

**field:** Example of convergence:  $a_n = \frac{1}{n}$

**field:** For any  $\epsilon > 0$ , choose  $N$  such that  $\frac{1}{N} < \epsilon$ . Then for any  $m > N$  we have that

$$a_n = \frac{1}{n} < \frac{1}{N} < \epsilon$$

and therefore  $|a_m - 0| = \frac{1}{n} < \epsilon$

**NOTE:**

**field:** 222

**field:** Given two convergent sequences  $\{a_n\}$  and  $\{b_m\}$  such that  $a_m \rightarrow a$  and  $b_m \rightarrow b$   
 $\lim_{n \rightarrow \infty} a_n b_n =$

**field:** Given two convergent sequences  $\{a_n\}$  and  $\{b_n\}$  such that  $a_n \rightarrow a$  and  $b_n \rightarrow b$   
 $\lim_{n \rightarrow \infty} a_n b_n = (\lim_{n \rightarrow \infty} a_n)(\lim_{n \rightarrow \infty} b_n) = ab$

**NOTE:**

**field:** 223

**field:** Definition: Convergence in probability

**field:** A sequence of random variables  $\{X_n\}_{n \geq 1}$  **converges in probability** to a random variable  $X$ , if for every  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0$$

We write  $X_n \xrightarrow{p} X$

Equivalently,  $x_n \xrightarrow{p} x$  if  $\lim_{n \rightarrow \infty} P(|x_n - x| < \epsilon) = 1$

**NOTE:**

**field:** 224

**field:** Convergence in probability example: Let  $\{x_n\}$  be a sequence of random variables such that  $x_n \sim N(0, 1/m^2)$

Show that  $x_n \xrightarrow{p} 0$ :

**field:** Let  $\epsilon > 0$ . We obtain  $P(|x_n - 0|) = P(x_n > \epsilon) + P(x_n < -\epsilon)$ . ie we are looking at the tail probabilities.

Now,

$$\begin{aligned} P(X_n < -\epsilon) + P(x_n > \epsilon) &= P(nx_n < n\epsilon) + P(nx_n > n\epsilon) \\ &= \Phi(n\epsilon) + 1 - \Phi(n\epsilon) \\ &= 2\Phi(-n\epsilon) \xrightarrow[n \rightarrow \infty]{} 0 \end{aligned}$$

Therefore  $x_n \xrightarrow{p} 0$

**NOTE:**

**field:** 225

**field:** Example convergence in probability Let  $W \sim N(0, 1)$  and  $U \sim Unif(0, 1)$ , and define the sequence  $\{x_n\}_{n \geq 1}$  as  $x_n = W$  with prob  $1 - 1/n$ ,  $U$  with prob  $1/n$

Show that  $x_n \xrightarrow{p} W$

**field:** Let  $\epsilon > 0$  Then.

$$\begin{aligned} P(|X_n - W| > \epsilon) &= P(|X_n - W| > \epsilon | X_n = W)P(X_n = W) \\ &\quad + P(|X_n - W| > \epsilon | X_n = U)P(X_n = U) \\ &= 0 \cdot (1 - 1/n) + p_n(1/n) \end{aligned}$$

Where  $p_n$  is a probability, and therefore  $0 \leq p_n \leq 1$

It follows that  $p_n \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$ , and therefore  $P(|X_n - W| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$ , for all  $\epsilon > 0$ , so that  $X_n \xrightarrow{p} W$ .

**NOTE:**

**field:** 226

**field:** Does  $X_n \xrightarrow{p} c$  imply  $E(X_n) \rightarrow c$ ?

**field:** Let  $X_n = 0$  with probability  $1 - 1/n$ ,  $n^2$  with probability  $1/n$  Then  $P(|X_n - 0| > \epsilon) \leq P(X_n = n^2) = 1/n \xrightarrow{n \rightarrow \infty} 0$  On the other hand,  $E(X_n) = 0 \cdot P(X_n = 0) + n^2 P(X_n = n^2) = 0 + n^2 \frac{1}{n} = n \xrightarrow{n \rightarrow \infty} \infty$ . Therefore  $X_n \xrightarrow{p} c$  does not imply  $E(X_n) \rightarrow c$

**NOTE:**

**field:** 227

**field:** Does  $E(X_n) \rightarrow c$  imply  $X_n \xrightarrow{p} c$ ?

**field:** No.

Let  $X_n = 0$ , with prob  $1 - 1/n$ ,  $n$  with prob  $1/n$ . Then  $E(X_n) = 0 \cdot P(X_n = 0) + nP(X_n = n) = 0 + n1/n = 1$  for all  $n$ . But  $P(|X_n - 0| > \epsilon) \leq P(X_n = n) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$ . It follows,  $X_n \xrightarrow{p} 0$ , and therefore we have  $E(X_n) \rightarrow c$  does not imply  $X_n \xrightarrow{p} c$

**NOTE:**

**field:** 228

**field:** Suppose  $\{X_n\}_{n \geq 1}$  and  $\{Y_n\}_{n \geq 1}$  be two sequences of random variables such that  $X_n \xrightarrow{p} x_0$  and  $Y_n \xrightarrow{p} y_0$  as  $n \rightarrow \infty$ , where  $x_0, y_0 \in \mathbb{R}$

What properties do we have?

**field:**

- $X_n \pm Y_n \xrightarrow{p} x_0 \pm y_0$  as  $n$  increases to  $\infty$
- $X_n Y_n \xrightarrow{p} x_0 y_0$  as  $n$  increases to  $\infty$
- $X_n / Y_n \xrightarrow{p} x_0 / y_0$  as  $n$  increases to infinity, provided that  $P(Y_n = 0) = 0$  for all  $n$  and  $y_0 \neq 0$

**NOTE:**

**field:** 229

**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that  $x_n \xrightarrow{p} x_0 \in \mathbb{R}$ , as  $n \rightarrow \infty$ , and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function. Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \rightarrow \infty$$

**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that  $x_n \xrightarrow{p} x_0 \in \mathbb{R}$ , as  $n \rightarrow \infty$ , and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \rightarrow \infty$$

**NOTE:**

**field:** 230

**field:** Proof of: Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that  $x_n \xrightarrow{p} x_0 \in \mathbb{R}$ , as  $n \rightarrow \infty$ , and let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a continuous function . Then

$$g(X_n) \xrightarrow{p} g(x_0) \text{ as } n \rightarrow \infty$$

**field:** Since  $g$  is continuous at  $X = x_0$ , we have that for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that  $|g(x) - g(x_0)| > \epsilon$  implies  $|x - x_0| > \delta$

We obtain

$$0 \leq P(|g(X_n) - g(x_0)| > \epsilon) \leq P(|X_n - x_0| > \delta) \xrightarrow{n \rightarrow \infty} 0$$

**NOTE:**

**field:** 231

**field:** Weak Law of Large numbers

**field:** Let  $X_1, X_2, X_3 \dots$  Be a sequence of iid random variables with  $E(X_1) = \mu$  (finite) and  $V(X_1) = \sigma^2 < \infty$ , and define  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  (the sample mean).

Then

$$\bar{X}_n \xrightarrow{p} \mu \text{ as } n \rightarrow \infty$$

**NOTE:**

**field:** 232

**field:** Proof of Weak Law of Large Numbers

**field:**

$$\begin{aligned} P(|\bar{X}_n - \mu| > \epsilon) &= P((\bar{X}_n - \mu)^2 > \epsilon^2) \\ &\leq \frac{E((\bar{X}_n - \mu)^2)}{\epsilon^2} \text{ by Chebyshev's Inequality} \\ &= \frac{V(\bar{X}_n)}{\epsilon^2} \text{ by def of variance} \\ &= \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Therefore  $\bar{X}_n \xrightarrow{p} \mu$

**NOTE:**

**field:** 233

**field:** Consistency

**field:** If our estimate converges in probability to the value of the parameter of interest as the sample size  $n$  increases

**NOTE:**

**field:** 234

**field:** Consistency of  $S^2$



**field:** Suppose  $X_1, X_2, \dots$  is a sequence of iid random variables with  $E(X_1) = \mu$  finite and  $V(X_1) = \sigma^2 < \infty$   
and define

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{The sample variance}$$

Can we show that  $S_n^2$  is a consistent estimate of  $\sigma^2$ ? In other words, can we show that  $S_n^2 \xrightarrow{p} \sigma^2$  as  $n \rightarrow \infty$

Using Chebychev's inequality, we obtain

$$\begin{aligned} P(|S_n^2 - \sigma^2| > \epsilon) &\leq \frac{E[(S_n^2 - \sigma^2)^2]}{\epsilon^2} \\ &= \frac{V(S_n^2)}{\epsilon^2} \end{aligned}$$

Therefore, a sufficient condition that  $S_n^2$  converges in probability to  $\sigma^2$  is that the variance of  $S_n^2$ ,  $V(S_n^2) \rightarrow 0$ , as  $n \rightarrow \infty$

**NOTE:**

**field:** <sup>235</sup>

**field:**  $V(S_n^2) \rightarrow 0$  as long as

**field:**  $V(S_n^2) \rightarrow 0$  as long as the fourth central moment  $\mu_4 = E[(X_1 - \mu)^4]$  is finite.

**NOTE:**

**field:** <sup>236</sup>

**field:** Khinchin's WLLN

**field:** Let  $X_1, X_2, \dots$  be a sequence of iid random variables with  $E(X_1) = \mu$  (finite). Then,  $\bar{X}_n \xrightarrow{p} \mu$  as  $n \rightarrow \infty$

**NOTE:**

**field:** 237

**field:** Let  $X_1, X_2, \dots$  be a sequence of random variables, such that for some  $r > 0$  and  $c \in \mathbb{R}$ ,  $E[|X_n - c|^r] \xrightarrow{n \rightarrow \infty} 0$ . Then  $X_n \xrightarrow{p} c$ , as  $n \rightarrow \infty$

**field:** (A general result to establish convergence in probability )

Let  $X_1, X_2, \dots$  be a sequence of random variables, such that for some  $r > 0$  and  $c \in \mathbb{R}$ ,  $E[|X_n - c|^r] \xrightarrow{n \rightarrow \infty} 0$ . Then  $X_n \xrightarrow{p} c$ , as  $n \rightarrow \infty$

**NOTE:**

**field:** 238

**field:** Consistent estimator for  $X_1, X_2, \dots, X_n \sim \text{iid Unif}(0, \theta)$ ,  $\theta > 0$ . ( and sketch of proof )

**field:**  $X_{(n)} = \max(X_1, \dots, X_n)$  (the largest order statistic)

Proof

First recall that the pdf of  $X_{(n)}$  is given by

$$f(x) = nx^{n-1}\theta^{-n}, 0 < x < \theta, 0 \text{ otherwise}$$

We obtain

$$\begin{aligned}
E(X_{(n)}) &= \int_0^\theta x f(x) dx \\
&= n\theta^{-n} \int_0^\theta x^n dx \\
&= \frac{n}{n-1} \theta \\
E(X_{(n)}^2) &= \int_0^\theta x^2 f(x) dx \\
&= n\theta^{-n} \int_0^\theta x^{n+1} dx \\
&= \frac{n}{n+2} \theta^2
\end{aligned}$$

We have

$$\begin{aligned}
E[(X_{(n)} - \theta)^2] &= E(X_{(n)}^2) - 2\theta E(X_{(n)}) + \theta^2 \\
&= \frac{n}{n+2} \theta^2 - 2\theta \frac{n}{n+1} \theta + \theta^2 \\
&\dots \\
&= \frac{2\theta^2}{(n+1)(n+2)} \xrightarrow{n \rightarrow \infty} 0
\end{aligned}$$

Hence, taking  $c = 0$  and  $r = 2$ , from the previous theorem, we obtain  $X_{(n)} \xrightarrow{P} \theta$  as  $n \rightarrow \infty$

**NOTE:**

**field:** 239

**field:** Definition Almost Sure Convergence

**field:** A sequence  $\{X_n\}_{n \geq 1}$  of random variables is said to converge **Almost Surely** to a random variable  $X$  if for every  $\epsilon > 0$ ,

$$P(\lim_{n \rightarrow \infty} |X_n - X| > \epsilon) = 0$$

We write  $X_n \xrightarrow{a.s} X$  as  $n \rightarrow \infty$

**NOTE:**

**field:** 240

**field:** Strong Law of Large Numbers

**field:** Let  $X_1, X_2, \dots$  be an iid sequence of random variables, with  $E(X_1) = \mu$  (finite) and  $V(X_1) = \sigma^2 < \infty$ . Then,

$$\bar{X}_n \xrightarrow{a.s.} \mu \quad \text{as } \mu \rightarrow \infty$$

**NOTE:**

**field:** 241

**field:** Does convergence in probability imply convergence almost surely?

**field:** No. Let  $\Omega = [0,1]$ , with uniform probability distribution. Define the sequence  $\{X_n\}_{n \geq 1}$  as:

$$\begin{aligned} X_1(\omega) &= \omega + \mathbb{I}_{[0,1]}(\omega) \\ X_2(\omega) &= \omega + \mathbb{I}_{[0,1/2]}(\omega) \\ X_3(\omega) &= \omega + \mathbb{I}_{[1/2,1]}(\omega) \\ X_4(\omega) &= \omega + \mathbb{I}_{[0,1/3]}(\omega) \\ X_5(\omega) &= \omega + \mathbb{I}_{[1/3,2/3]}(\omega) \\ &\vdots \end{aligned}$$

$$X_5(\omega) = \omega + 1$$

Let  $X(\omega) = \omega$ , then it is easy to show that  $X_n \xrightarrow{P} X$  because  $P(|X_n - X| \geq \epsilon) = P([a_n, b_n])$ , where  $l_n = \text{length}([a_n, b_n]) \xrightarrow{n \rightarrow \infty} 0$ .

However  $X_n$  does not converge to  $X$  almost surely, because for every  $\omega \in [0,1]$ , alternates between  $\omega$  and  $\omega + 1$ , infinitely often as  $n \rightarrow \infty$

**NOTE:**

**field:** 242

**field:** Convergence in Distribution

**field:** A sequence  $\{X_n\}_{n \geq 1}$  of random variables converges in distribution to a random variable  $X$  if,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

at all points  $x$  where  $F_X(x)$  is continuous

We write  $X_n \xrightarrow{d} X$

**NOTE:**

**field:** 243

**field:** Example of convergence in distribution

Let  $X_n \sim N(0, \frac{n+1}{n})$ , and  $X \sim N(0, 1)$ . We want to show that  $X_n \xrightarrow{d} X$ .

**field:**

$$\begin{aligned} P(X_n \leq X) &= P\left(\sqrt{\frac{n}{n+1}} X_n \leq \sqrt{\frac{n}{n+1}} x\right) \\ &= \Phi\left(\sqrt{\frac{n}{n+1}} x\right) \xrightarrow{n \rightarrow \infty} \Phi(x) \end{aligned}$$

And we obtain that  $F_{X_n} \rightarrow \Phi(x) = F_X(x), \forall x$ , and therefore  $X_n \xrightarrow{d} X$

**NOTE:**

**field:** 244

**field:** Does Convergence in probability imply convergence in distribution?

**field:** Yes

**NOTE:**

**field:** <sup>245</sup>

**field:** Does Convergence in distribution imply convergence in probability?

**field:** No - unless converges in distribution to a constant

**NOTE:**

**field:** <sup>246</sup>

**field:** A sequence  $\{X_n\}_{n \geq 1}$  of random variables converges in probability to a constant  $c \in \mathbb{R}$  if and only if

**field:** A sequence  $\{X_n\}_{n \geq 1}$  of random variables converges in probability to a constant  $c \in \mathbb{R}$  if and only if the sequence converges in distribution to  $c$

**NOTE:**

**field:** <sup>247</sup>

**field:** If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  we have that

1.  $X_n \pm Y_n$
2.  $X_n Y_n$

**field:** In general it is not true that if  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  we have that

1.  $X_n \pm Y_n \xrightarrow{d} X + Y$
2.  $X_n Y_n \xrightarrow{d} XY$

**NOTE:**

**field:** <sup>248</sup>

**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that  $X_n \xrightarrow{d} X$ , for some random variable  $X$  (possibly a constant). Then for any continuous function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have  $g(X_n) \xrightarrow{d}$

**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that  $X_n \xrightarrow{d} X$ , for some random variable  $X$  (possibly a constant). Then for any continuous function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we have  $g(X_n) \xrightarrow{d} g(X)$

**NOTE:**

**field:** 249

**field:** Let  $\{X_n\}_{n \geq 1}$  and  $\{Y_n\}_{n \geq 1}$  be two sequences of random variables such that  $X_n \xrightarrow{d} X$  for some random variable  $X$  (possibly a constant) and  $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as  $n \rightarrow \infty$ ,

1.  $X_n \pm Y_n \xrightarrow{d}$
2.  $X_n Y_n \xrightarrow{d}$
3.  $X_n / Y_n \xrightarrow{d}$  provided  $P(Y_n = 0) = 0 \forall n$  and  $c \neq 0$

**field:** Slutsky's Theorem Let  $\{X_n\}_{n \geq 1}$  and  $\{Y_n\}_{n \geq 1}$  be two sequences of random variables such that  $X_n \xrightarrow{d} X$  for some random variable  $X$  (possibly a constant) and  $Y_n \xrightarrow{p} c \in \mathbb{R}$

Then, as  $n \rightarrow \infty$ ,

1.  $X_n \pm Y_n \xrightarrow{d} X \pm c$
2.  $X_n Y_n \xrightarrow{d} cX$
3.  $X_n / Y_n \xrightarrow{d} X/c$  provided  $P(Y_n = 0) = 0 \forall n$  and  $c \neq 0$

**NOTE:**

**field:** 250

**field:** Central Limit Theorem

**field:** Let  $X_1, X_2, \dots$  be an iid sequence of random variables, with  $E(X_1) = \mu$  (finite) and  $V(X_1) = \sigma^2 < \infty$

Then, for  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  (the sample mean), we have that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1) \quad \text{as } n \rightarrow \infty$$

**NOTE:**

**field:** 251

**field:** Equivalent results of CLT

**field:**

- $\frac{(\bar{X}_n - \mu)}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$
- $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$
- $\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1)$
- $\bar{X}_n \xrightarrow{d} N(\mu, \sigma^2/n)$

**NOTE:**

**field:** 252

**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that the mgf  $M_{X_n}(t)$  of  $X_n$  exists in a neighborhood of 0, for all  $n$ , and suppose that

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where  $M_X(t)$  is the mgf for some random variable  $X$ . Then,



**field:** Let  $\{X_n\}_{n \geq 1}$  be a sequence of random variables such that the mgf  $M_{X_n}(t)$  of  $X_n$  exists in a neighborhood of 0, for all  $n$ , and suppose that

$$\lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t) \quad \text{for all } t \text{ in a neighborhood of } 0$$

where  $M_X(t)$  is the mgf for some random variable  $X$ . Then, there exists a unique cdf  $F_X(x)$  whose moments are determined by  $M_X(t)$  and for all  $x$ , where  $F_X(x)$  is continuous we have  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$

**NOTE:**

**field:** 253

**field:**  $\frac{\sqrt{n}(\bar{X} - \mu)}{S_n} \xrightarrow{d}$

**field:** Using the CLT, and Slutsky's theorem, we have

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \frac{\sigma}{S_n} \cdot \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

**NOTE:**

**field:** 254

**field:**  $g(X) \approx$   
 $E(g(X)) \approx, V(g(X)) \approx$

**field:**

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu)$$

Using a first order Taylor approximation  $E(g(X)) \approx g(\mu), V(g(X)) \approx [g'(\mu)]^2 V(X)$

**NOTE:**

**field:** 255

**field:** Delta Method

**field:** Let  $\{Y_n\}_{n \geq 1}$  be a sequence of random variables such that  $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$  as  $n \rightarrow \infty$ . Suppose that for a given function  $g$  and a specific value of  $\theta$ ,  $g'(\theta)$  exists and is not equal to zero. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2 [g'(\theta)]^2)$$

as  $n \rightarrow \infty$

**NOTE:**

**field:** 256

**field:** Second Order delta method

**field:** Let  $\{Y_n\}_{n \geq 1}$  be a sequence of random variables such that  $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$  as  $n \rightarrow \infty$ . And that for a given function  $g$  as specific value of  $\theta$ , we have  $g'(\theta) = 0$ , but  $g''(\theta)$  exists and is not equal to 0. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi_1^2 \quad \text{as } n \rightarrow \infty$$

**NOTE:**

**field:** 257

**field:**  $\chi_n^2 \sim$  for sufficiently large  $n$

**field:**  $\chi_n^2 \sim N(n, 2n)$

**NOTE:**

**field:** 258

**field:** Definition Statistic

**field:** Let  $X_1, \dots, X_n$  be a random sample from a given population. Then, any observable real-valued (or vector-valued) function  $T(\mathbf{X}) = T(X_1, \dots, X_n)$  of the random variables  $X_1, \dots, X_n$  is called a **Statistic**

**NOTE:**

**field:** 259

**field:** Sampling Distribution

**field:** The probability distribution of the statistic  $T(\mathbf{X})$  is called the **Sampling Distribution** of  $T(\mathbf{X})$

**NOTE:**

**field:** 260

**field:** Sufficient Statistic

**field:** A statistic  $T(\mathbf{X})$  is a **Sufficient Statistic** for  $\theta$ , if the conditional distribution of the sample  $\mathbf{X}$  given the value of  $T(\mathbf{X})$  does not depend on  $\theta$

**NOTE:**

**field:** 261

**field:** Determine if  $T(\mathbf{X}) = \sum X_i$  where  $X_i \sim \text{Bern}(p)$  is sufficient for  $p$  using definition of sufficiency

**field:**

$$\begin{aligned}
 P(\mathbf{X} = \mathbf{x} | T = t) &= \frac{P(\cap_{i=1}^n X_i = x_i)}{P(T = t)} \\
 &= \prod_{i=1}^n \frac{P(X_i = x_i)}{P(T = t)} \quad \text{by independence} \\
 &= \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{Because } T \sim \text{Binom}(n, p) \\
 &= \frac{p^t (1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \quad \text{because } t = \sum_{i=1}^n x_i \\
 &= \frac{1}{\binom{n}{t}} \quad \text{which is free of } p
 \end{aligned}$$

**NOTE:**

**field:** 262

**field:** How to show sufficiency (not using factorization)

**field:** Let  $p(\mathbf{X}|\theta)$  be the joint PDF or PMF of  $\mathbf{X}$  and  $q(t|\theta)$  the PDF or PMF of the statistic  $T(\mathbf{X})$ . Then  $T(\mathbf{X})$  is a sufficient statistic for  $\theta$  if for every  $\mathbf{X}$  in the sample space, the ratio

$$\frac{p(\mathbf{x}|\theta)}{q(T(\mathbf{x})|\theta)}$$

is constant as a function of  $\theta$

**NOTE:**

**field:** 263

**field:** Suppose that  $X_1, \dots, X_n$  are iid  $N(\mu, \sigma^2)$  where  $\sigma^2$  is known. If the statistic  $T(\mathbf{X}) = \bar{X}_n$  sufficient for  $\mu$ ?

**field:**

$$\begin{aligned}\frac{f(\mathbf{x}|\mu)}{q(T(\mathbf{X})|\mu)} &= \frac{(2\pi\sigma^2)^{n/2}e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^n(x_i-\bar{x})^2+n(\bar{x}-\mu)^2]}}{(2\pi\sigma/n)^{-1/2}e^{-\frac{1}{2\sigma^2}(\bar{x}-\mu)^2}} \\ &= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\bar{x})^2}\end{aligned}$$

Which does not depend on  $\mu$ , and therefore  $\bar{X}_n$  is sufficient for  $\mu$  as long as  $\sigma^2$  is known

**NOTE:**

**field:** 264

**field:** The joint pdf of the sample  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is Suppose that  $X_1, \dots, X_n$  are iid  $N(\mu, \sigma^2)$  where  $\sigma^2$  is known.

**field:**

$$\begin{aligned}f(\mathbf{x}|\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\ &= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\mu)^2} \\ &= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x}+\bar{x}-\mu)^2} \\ &= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 \sum_{i=1}^n (x_i-\bar{x})^2 + 2(\bar{x}-\mu) \sum_{i=1}^n (x_i-\bar{x}) + n(\bar{x}-\mu)^2} \\ &= (2\pi\sigma^2)^{n/2} e^{-1/2\sigma^2 (\sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\mu)^2)}\end{aligned}$$

**NOTE:**

**field:** 265

**field:** Show a statistic  $T(\mathbf{X})$  is sufficient

**field:** Neyman factorization theorem Let  $f(\mathbf{x}|\theta)$  denote the joint pdf or pmf of the sample  $\mathbf{X}$ , A statistic  $T(\mathbf{X})$  is a sufficient statistic for  $\theta$  if and only if there exists functions  $g(t|\theta)$  and  $h(\mathbf{x})$  such that for all sample points  $\mathbf{x}$  and all values of  $\theta$  we can write

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$$

Note, in the theorem

- The function  $g(T(\mathbf{X})|\theta)$  depends on  $\mathbf{x} = (x_1, \dots, x_n)$  only through the statistic  $T(\mathbf{X})$ .
- The function  $h(\mathbf{X})$  does not depend on  $\theta$

**NOTE:**

**field:** 266

**field:** Exponential Family

**field:**

$$f(\mathbf{X}|\theta) = \mathbf{h}(\mathbf{x})\mathbf{c}(\theta)e^{\sum_{i=1}^n \mathbf{w}_i((\theta))\mathbf{t}_i(\mathbf{x})}$$

**NOTE:**

**field:** 267

**field:** Sufficiency in the exponential family

**field:** Let  $X_1, \dots, X_n$  be iid observations from a PDF or PMF,  $f(x|\boldsymbol{\theta})$  that belongs to an exponential family of the form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})e^{\sum_{i=1}^k w_i(\boldsymbol{\theta})t_i(x)}$$

Where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$ ,  $d \leq k$ . Then

$$T(\mathbf{X}) = \left( \sum_{j=1}^k t_1(x_j), \dots, \sum_{j=1}^k t_k(x_j) \right)$$

**NOTE:**

**field:** 268

**field:** Minimal Sufficient Statistic

**field:** A sufficient statistic  $T(\mathbf{X})$  is called a **Minimal Sufficient Statistic** if for any other sufficient statistic  $T'(\mathbf{X})$ ,  $T(\mathbf{X})$  is a function of  $T'(\mathbf{X})$

**NOTE:**

**field:** 269

**field:** Determining if a statistic is minimal sufficient

**field:** Let  $f(x|\theta)$  be the PDF or PMF of a sample  $\mathbf{X}$ . Suppose there exists a function  $T(x)$  such that, for every two sample points,  $\mathbf{x}$  and  $\mathbf{y}$ , the ratio  $\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$  is constant as a function of  $\theta$  iff and only if  $T(\mathbf{x}) = T(\mathbf{y})$ . Then  $T(\mathbf{x})$  is a minimal sufficient statistic for  $\theta$ .

**NOTE:**

**field:** 270

**field:** Example of finding a minimal sufficient statistic: Suppose that  $X_1, \dots, X_n$  are iid Bernoulli( $p$ ). What is a minimal sufficient statistic for  $p$ ?

**field:**

$$\begin{aligned} f(\mathbf{x}|p) &= \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

And therefore for any two sample points  $\mathbf{x}$  and  $\mathbf{y}$ , we obtain

$$\begin{aligned}\frac{f(\mathbf{x}|p)}{f(\mathbf{y}|p)} &= \frac{p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i}}{p^{\sum_{i=1}^n y_i} (1-p)^{n-\sum_{i=1}^n y_i}} \\ &= p^{\sum_{i=1}^n x_i - \sum_{i=1}^n y_i} (1-p)^{\sum_{i=1}^n y_i - \sum_{i=1}^n x_i}\end{aligned}$$

Which is constant as a function of  $p$  iff  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

Hence it follows from Lehman-Sheffe that  $T(\mathbf{x}) = \sum_{i=1}^n x_i$  is minimal sufficient for  $p$

**NOTE:**

**field:** 271

**field:** Minimal sufficient statistic for  $\mu, \sigma^2$ , where the  $X$ s are  $N(\mu, \sigma^2)$

**field:**  $T(\mathbf{x}) = (\bar{x}, S_x^2)$  by Lehmann-Schaffe is minimal sufficient.

**NOTE:**

**field:** 272

**field:** Facts about sufficiency

**field:**

- The entire sample  $\mathbf{X}$  is always sufficeint.
- Any one-to-one funciton of a minimal sufficient statisitc is also a minimal sufficient statistic

**NOTE:**

**field:** 273

**field:** Ancillary Statistic



**field:** A statistic  $S(\mathbf{X})$  whose distribution does not depend on the parameter  $\theta$  is called an ancillary statistic for  $\theta$

**NOTE:**

**field:** <sub>274</sub>

**field:** Complete statistic

**field:** Let  $f(t|\theta)$  be the family of pdf's or pmfs for a statistic  $T = T(\mathbf{x})$ .

The family of probability distributions is called **complete** (with respect to  $\theta$ ) if  $E_\theta(g(t)) = 0$  for all  $\theta$ , implies  $P_\theta(g(T) = 0) = 1$  for all  $\theta$

Equivalently, we say that  $T = T(\mathbf{X})$  is a complete statistic.

In short, a statistic  $T = T(\mathbf{x})$  is complete, if  $E_\theta(g(T)) = 0$  for all  $\theta$  implies  $g(t) = 0$  with probability 1

**NOTE:**

**field:** <sub>275</sub>

**field:** (Binomial complete sufficient statistic)

**field:** Suppose the statistic  $T \sim \text{Binom}(n, p)$ ,  $0 < p < 1$ , and let  $g$  be a function such that  $E_p(g(T)) = 0$  for all  $p$ .

Then, with  $r = (\frac{p}{1-p})^t$

$$\begin{aligned}
0 &= E_p(g(T)) \\
&= \sum_{t=0}^n g(t) \binom{n}{t} p^t (1-p)^{n-t} \\
&= (1-p)^n \sum_{t=0}^n g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t \\
&= (1-p)^n \sum_{t=0}^n g(t) \binom{n}{t} r^t \\
&\neq 0. \text{ This is a polynomial of degree } n \text{ in } r \text{ with coefficients } g(t) \binom{n}{t}
\end{aligned}$$

For the polynomial to be 0 for all  $r$  (and consequently for all  $p$ ) each coefficient must be zero and therefore it must be the case that  $g(t) = 0$  for  $t = 0, 1, 2, \dots, n$ . Since  $T \sim \text{Binom}(n, p)$ , we have that  $T$  takes on the values  $t = 0, 1, 2, \dots, n$  with probability 1 and therefore, we obtain  $P_p(g(T) = 0) = 1$ . Hence  $T$  is a complete statistic.

**NOTE:**

**field:**  $_{276}$

**field:**  $\text{Uniform}(0, \theta)$  complete sufficient statistic

**field:**  $T(\mathbf{X}) = X_{(n)}$

Suppose that  $X_1, \dots, X_n$  are iid  $\text{Uniform}(0, \theta)$ ,  $\theta > 0$ . We know that  $T(\mathbf{X}) = X_{(n)}$  (the max order statistic) is sufficient for  $\theta$ . Furthermore ,

$$f(t|\theta) = nt^{n-1}\theta^{-n} \quad 0 < t < \theta$$

Now suppose that  $g(t)$  is a function satisfying  $E_\theta(g(T)) = 0, \forall \theta$ . Differentiating on both sides with respect to  $\theta$ ,

$$\begin{aligned}
0 &= \frac{d}{d\theta} E_\theta(g(t)) \\
&= \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} \theta^{-n} dt \\
&= \theta^{-n} \frac{d}{d\theta} \int_0^\theta g(t) n t^{n-1} dt + \left( \frac{d}{d\theta} \theta^{-n} \right) \int_0^\theta g(t) n t^{n-1} dt \\
&= \theta^{-n} g(\theta) n \theta^{n-1} + 0
\end{aligned}$$

Since  $n\theta^{-1} \neq 0$ , we must have that  $g(\theta) = 0 \quad \forall \theta > 0$ . And therefore  $T$  is complete.

**NOTE:**

**field:** 277

**field:** Does minimal sufficient imply complete?

**field:** No

Suppose that  $X_1, \dots, X_n$  are iid  $N(\theta, \theta^2)$  where  $\theta \in \mathbb{R}$  is the unknown parameter of interest.

We have

$$\begin{aligned}
\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} &= \frac{(2\phi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2}}{(2\phi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2} [\sum_{i=1}^n x_i^2 - 2\theta \sum_{i=1}^n x_i]}}{e^{-\frac{1}{2\sigma^2} [\sum_{i=1}^n y_i^2 - 2\theta \sum_{i=1}^n y_i]}}
\end{aligned}$$

Which is free of  $\theta$  if  $\sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i^2$  and  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$

It follows that  $T(\mathbf{X}) = (\sum_{i=1}^n x_i, \sum_{i=1}^n x_i^2)$  is minimal sufficient for  $\theta$

Now observe that  $T_1(\mathbf{X}) = \sum_{i=1}^n x_i \sim N(n\theta, n\theta^2)$  and therefore

$$\begin{aligned}
E(T_1^2) &= V(T_1) + [E(T_1)]^2 \\
&= n\theta^2 + n^2\theta^2 \\
&= n\theta^2(1 + n)
\end{aligned}$$

On the other hand, for  $T_2 = \sum_{i=1}^n x_i^2$ ,

$$\begin{aligned} E(T_2) &= nE(X_1)^2 \\ &= n[V(X_1) + [E(X_1)]^2] \\ &= n\theta^2 + n\theta^2 \\ &= 2n\theta^2 \end{aligned}$$

Then, taking  $h(t_1, t_2) = 2t_1^2 - (n+1)t_2$ , we have

$$\begin{aligned} E_\theta[h(T_1, T_2)] &= E_\theta[2T_1^2 - (n+1)T_2] \\ &= 2E_\theta(T_1^2) - (n+1)E(T_2) \\ &= 2n(n+1)\theta^2 - 2n(n+1)\theta^2 \\ &= 0 \quad \forall \theta \end{aligned}$$

But because  $h(\mathbf{t}) \neq 0 \quad \forall \theta$ , we have that  $T(\mathbf{X})$  is not complete.

**NOTE:**

**field:** 278

**field:** Complete statistics in the exponential family

**field:** Let  $X_1, \dots, X_n$  be iid observations from an exponential family. with PDF or PMF of the form

$$f(x|\theta) = h(x)c(\theta)e^{\sum_{j=1}^k \omega_j(\theta_j)t_j(x)}$$

Where  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$

Then, the statistic  $T(\mathbf{X}) = (\sum_{i=1}^n t_1(x_i), \sum_{i=1}^n t_2(x_i), \dots, \sum_{i=1}^n t_k(x_i))$  is complete, as long as the parameter space  $\Theta$  contains an open set in  $\mathbb{R}^k$

**NOTE:**

**field:** 279

**field:** Suppose that a statistic  $T$  is complete and let  $g$  be a one-to-one function. Is the statistic  $U = g(T)$  also complete?

**field:** Yes

**NOTE:**

**field:** <sub>280</sub>

**field:** Does complete statistic imply minimal sufficient statistic?

**field:** If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic

**NOTE:**

**field:** <sub>281</sub>

**field:** Basu's Theorem

**field:** If  $T(\mathbf{x})$  is a complete and minimal sufficient statistic, then  $T(\mathbf{x})$  is an independent of every ancillary statistic.

**NOTE:**

**field:** <sub>282</sub>

**field:** Likelihood function

**field:** Let  $f(\mathbf{x}|\theta)$  denote the joint pdf or pmf of the sample  $\mathbf{X} = (X_1, \dots, X_n)$ , then given that  $\mathbf{X} = \mathbf{x}$  is observed, the function of  $\theta$  defined as

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

is called the Likelihood Function

**NOTE:**

**field:** <sub>283</sub>

**field:** Idea of likelihood function

**field:** Suppose that  $\mathbf{X}$  is a discrete random vector (so we can interpret probabilities easier)

Then  $L(\theta|\mathbf{x}) = P_\theta(\mathbf{X} = \mathbf{x})$ . Now if we compare the likelihood function at two parameter values  $\theta_1, \theta_2$  and we observe that

$$P_{\theta_1}(\mathbf{X} = \mathbf{x}) = L(\theta_1|\mathbf{x}) > L(\theta_2|\mathbf{x}) = P_{\theta_2}(\mathbf{X} = \mathbf{x})$$

Then, the sample point  $\mathbf{x}$  that we actually observed is more likely to have occurred if  $\theta = \theta_1$ , than if  $\theta = \theta_2$ , which can be interpreted as that  $\theta_1$ , is a more plausible value for the true value of  $\theta$  than  $\theta_2$  is.

**NOTE:**

**field:** 284

**field:** Fisher information - one parameter case

**field:** Let  $X$  be a random variable with pdf or pmf  $f(x|\theta)$  where  $\theta \in \Theta \subseteq \mathbb{R}$  (Fisher ) information about  $\theta$  contained in  $X$  is

$$I_X(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta} \log f(x|\theta)\right)^2\right]$$

**NOTE:**

**field:** 285

**field:** Example of one parameter case Fisher information Suppose that  $X \sim \text{Bern}(p)$  What is the information that  $X$  contains about the parameter  $p$ ?

**field:** We have that  $f(x|p) = p^x(1-p)^{1-x}$ . Then

$$\log f(x|p) = x \log p + (1-x) \log(1-p)$$

$$\frac{\partial}{\partial p} \log f(x|p) = \frac{x}{p} - \frac{1-x}{1-p}$$

We obtain

$$\begin{aligned} \left(\frac{\partial}{\partial p} \log f(x|p)\right)^2 &= \left(\frac{x}{p} - \frac{1-x}{1-p}\right)^2 \\ &= \frac{x^2}{p^2} - \frac{2x(1-x)}{p(1-p)} + \frac{(1-x)^2}{(1-p)^2} \\ &= \frac{x^2}{p^2} - \frac{2(x-x^2)}{p(1-p)} + \frac{(1-2x+x^2)}{(1-p)^2} \end{aligned}$$

Therefore,

$$\begin{aligned} I_x(p) &= E_p\left[\left(\frac{\partial}{\partial p} \log f(x|p)\right)^2\right] \\ &= \frac{p}{p^2} - \frac{2(p-p)}{p(1-p)} + \frac{1-2p+p}{(1-p)^2} \\ &= \frac{1}{p} + \frac{1}{1-p} \\ &= \frac{1}{p(1-p)} \end{aligned}$$

**NOTE:**

**field:** <sub>286</sub>

**field:**

$$I_x(\theta) = E_\theta\left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta)\right)^2\right] =$$

**field:** If  $f(x|\theta)$  satisfies

$$\frac{\partial}{\partial \theta} E_{\theta} \left( \frac{\partial}{\partial \theta} \log f(x|\theta) \right) = \int \frac{\partial}{\partial \theta} \left[ \frac{\partial}{\partial \theta} \log f(x|\theta) \right] f(x|\theta) dx$$

$$I_x(\theta) = E_{\theta} \left[ \left( \frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 \right] = -E_{\theta} \left( \frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right)$$

**NOTE:**

**field:** 287

**field:** Suppose that  $X_1, \dots, X_n$  are iid observations with common pdf or pmf  $f(x|\theta)$ . Then, the information about  $\theta$  contained in the sample  $\mathbf{X} = (X_1, \dots, X_n)$  is

**field:**

$$I_{\mathbf{X}}(\theta) = nI_{X_1}(\theta)$$

**NOTE:**

**field:** 288

**field:** Fisher Information - multiparameter case

**field:** Let  $X$  be a random variable with pdf or pmf  $f(x|\boldsymbol{\theta})$ , where  $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2$ . Denote by

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[ \left( \frac{\partial}{\partial \theta_i} \log f(x|\boldsymbol{\theta}) \right) \left( \frac{\partial}{\partial \theta_j} \log f(x|\boldsymbol{\theta}) \right) \right] = -E_{\boldsymbol{\theta}} \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x|\boldsymbol{\theta}) \right]$$

For  $i, j = 1, 2$ . Then the (fisher) information matrix about  $\boldsymbol{\theta}$  is

$$I_x(\boldsymbol{\theta}) = \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) \end{pmatrix}$$



**NOTE:**

**field:** 289

**field:** Find Fisher information for Normal RVs

**field:** We have that  $\boldsymbol{\theta} = (\mu, \sigma^2)$  and  $f(x|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$   
Then,

$$\frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}) = \frac{\partial}{\partial} \left[ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2 \right] = \frac{(x-\mu)}{\sigma^2}$$

$$\frac{\partial}{\partial \sigma^2} = \frac{1}{2\sigma^2} \left[ \frac{(x-\mu)^2}{\sigma^2} - 1 \right]$$

$$\text{Therefore } I_{11} = E_{\theta} \left[ \left( \frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}) \right)^2 \right] = E_{\theta} \left[ \frac{(x-\mu)^2}{\sigma^4} \right] = \frac{1}{\sigma^4} \sigma^2 = \frac{1}{\sigma^2}$$

$$\begin{aligned} I_{22}(\boldsymbol{\theta}) &= E_{\theta} \left[ \frac{\partial}{\partial \sigma^2} \log f(x|\boldsymbol{\theta})^2 \right] \\ &= E_{\theta} \left\{ \left[ \frac{1}{2\sigma^2} \left( \frac{(x-\mu)^2}{\sigma^2} - 1 \right) \right]^2 \right\} \\ &= \frac{1}{4\sigma^4} E_{\theta} \left[ \left( \frac{(x-\mu)^2}{\sigma^2} - 1 \right)^2 \right] \\ &= \frac{1}{4\sigma^4 \cdot 2} \\ &= \frac{1}{2\sigma^4} \quad \text{Since } = V(\chi_1^2) \end{aligned}$$

Now for the off diagonal elements,

$$\begin{aligned} I_{12}(\boldsymbol{\theta}) &= I_{21}(\boldsymbol{\theta}) = E_{\theta} \left[ \left( \frac{\partial}{\partial \mu} \log f(x|\boldsymbol{\theta}) \right) \left( \frac{\partial}{\partial \sigma^2} \log f(x|\boldsymbol{\theta}) \right) \right] \\ &= E_{\theta} \left[ \frac{(x-\mu)}{\sigma^2} \frac{1}{2\sigma^2} \left[ \frac{x-\mu}{\sigma^2} \cdot 1 \right] \right] \\ &= \frac{1}{2\sigma^4} E_{\theta} \left[ \frac{(x-\mu)^3}{\sigma^3} - (x-\mu) \right] \end{aligned}$$

But  $E_{\theta}[(x-\mu)^3] = E_{\theta}[(x-\mu)] = 0$ , because  $X$  is symmetric around  $\mu$ ,  
and we obtain  $I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0$

We obtain that

$$\begin{aligned} I_{x_1}(\boldsymbol{\theta}) &= \begin{pmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta}) & I_{22}(\boldsymbol{\theta}) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \end{aligned}$$

And hence

$$I_{\mathbf{x}}(\boldsymbol{\theta}) = nI_{X_1}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

**NOTE:**

**field:** 290

**field:**  $I_T(\theta) \leq$

**field:**  $I_T(\theta) \leq I_{\mathbf{X}}(\theta)$  (The information of the statistic is less than or equal to the information of the sample)

**NOTE:**

**field:** 291

**field:** Let  $\mathbf{X} = X_1, \dots, X_n$  denote the entire data, and let  $T = T(\mathbf{X})$  be some statistic. Then, for all  $\theta \in \Theta \subseteq \mathbb{R}$ ,  $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$  Where the equality is attained...

**field:** Let  $\mathbf{X} = X_1, \dots, X_n$  denote the entire data, and let  $T = T(\mathbf{X})$  be some statistic. Then, for all  $\theta \in \Theta \subseteq \mathbb{R}$ ,  $I_{\mathbf{X}}(\theta) \geq I_t(\theta)$  Where the equality is attained if and only iff  $T(\mathbf{X})$  is sufficient for  $\theta$

**NOTE:**

**field:** 292

**field:** Let  $\mathbf{X} = (X_1, \dots, X_n)$ , denote a sample of iid observations and suppose the statistic  $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$  is such that  $T_1$  and  $T_2$  are independent. Then

$$I_T(\boldsymbol{\theta}) =$$

**field:** Let  $\mathbf{X} = (X_1, \dots, X_n)$ , denote a sample of iid observations and suppose the statistic  $T(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}))$  is such that  $T_1$  and  $T_2$  are independent. Then

$$I_T(\boldsymbol{\theta}) = I_{T_1}(\boldsymbol{\theta}) + I_{T_2}(\boldsymbol{\theta})$$

**NOTE:**

**field:** <sub>293</sub>

**field:** Point estimator

**field:** Any statistic  $T(\mathbf{X})$  that is used to estimate the value of a parameter is called a point estimator of  $\theta$ . We write  $\hat{\theta} = T(\mathbf{X})$

**NOTE:**

**field:** <sub>294</sub>

**field:** Method of moments

**field:**

$$\begin{aligned} m_1 &= \frac{1}{n} \sum_{i=1}^n X_i^1, & \mu_1 &= E(X^1) \\ m_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2, & \mu_2 &= E(X^2) \\ &\vdots \\ m_k &= \frac{1}{n} \sum_{i=1}^n X_i^k & \mu_k &= E(X^k) \end{aligned}$$

Equating and solving for  $\theta$  gives the MoM estimators

**NOTE:**

**field:** 295

**field:** Example Method of Moments Suppose that  $X_1, \dots, X_n$  are iid Binomial( $k, p$ ), where both  $k$  and  $p$  are unknown.

**field:** We have that

$$P(X_i = x|k, p) = \binom{k}{x} p^x (1-p)^{k-x}, x = 0, 1, \dots, k$$

and we obtain  $E(X_1) = kp$ ,  $E(X_1^2) = kp(1-p) + k^2p^2$

Solving the sytem of equations we obtain

$$\begin{aligned} m_1 &= \frac{1}{n} \sum_{i=1}^n X_i = kp \\ m_2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 = kp(1-p) + k^2p^2 \end{aligned}$$

Sovling the system of equations:

$$\tilde{p} = \frac{\bar{x}}{\tilde{k}}$$

$$\tilde{k} = \frac{\bar{x}^2}{\bar{x} - \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Possible problems:  $k$  has to be an integer, and not negative. (Estimates of parameters that are outside of the parameter space. )

**NOTE:**

**field:** 296

**field:** Maximum Likelihood Estimator

**field:** In this context, we define the **Maximum Likelihood Estimator (MLE)** of  $\theta$  as the parameter value  $\hat{\theta}_{ML} = \hat{\theta}(\mathbf{x})$  that satisfies

$$L(\hat{\theta}_{ML}|\mathbf{x}) = \sup_{\theta \in \Theta} L(\theta|\mathbf{x})$$

Note this often proceeds as taking the derivative of the log likelihood function and setting to zero to solve for parameters - not always

**NOTE:**

**field:** 297

**field:** Example of MLE Suppose that  $X_1, \dots, X_n$  are iid Exponential( $\lambda$ ). Find the MLE  $\hat{\lambda}_{ML}$  of  $\lambda$

**field:** Suppose that  $X_1, \dots, X_n$  are iid Exponential( $\lambda$ ). Find the MLE  $\hat{\lambda}_{ML}$  of  $\lambda$

We have that  $f(x|\lambda) = \frac{1}{\lambda}e^{-x/\lambda}$ ,  $x > 0$ , and therefore

$$L(\lambda|x) = \prod_{i=1}^n \frac{1}{\lambda} e^{-x_i/\lambda} = \lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^n x_i}$$

Since  $\log(\cdot)$  is a strictly monotone (one-to-one) and increasing, we consider instead the maximization of the log-likelihood

$$l(\lambda|\mathbf{x}) = \log L(\lambda|\mathbf{x}) = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^n x_i$$

$$\frac{\partial}{\partial \lambda} l(\lambda|\mathbf{x}) = \frac{-n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^n x_i$$

Solving  $\frac{\partial}{\partial \lambda} l(\lambda|\mathbf{x}) = 0$ , we obtain

$$\frac{-n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^n x_i = 0$$

$$-n\lambda + n\bar{x} = 0$$

$$\lambda = \bar{x}$$

**NOTE:**

**field:** 298

**field:** Example of MLE when can't differentiate

Suppose that  $X_1, \dots, X_n$  are iid Uniform( $0, \theta$ ),  $\theta > 0$ . Find the MLE of  $\theta$

**field:** We have that  $f(x|\theta) = \frac{1}{\theta}I(0 < x < \theta)$

And therefore

$$\begin{aligned} L(\theta|\mathbf{x}) &= \prod_{i=1}^n \frac{1}{\theta} I(0 < x_i < \theta) \\ &= \frac{1}{\theta^n} I(X_{(1)} > 0) I(X_{(n)} < \theta) \end{aligned}$$

In this case, the support of  $X$  depends on  $\theta$  and the maximization problem only makes sense whenever  $L(\theta|\mathbf{x}) > 0$ . We cannot simply approach the problem by taking partial derivatives, but assuming the likelihood is positive, we notice that  $L(\theta|\mathbf{x})$  is decreasing as a function of  $\theta$ , for  $\theta > X_{(n)}$

Picture with  $L(\theta)$  as zero until  $X_{(n)}$  on x axis, goes up to  $1/X_{(n)}$  there and decreases with  $\frac{1}{\theta^n}$

It follows the MLE of  $\theta$  is  $\hat{\theta}_{ML} = X_{(n)}$

**NOTE:**

**field:** 299

**field:** If  $\hat{\theta}_{ML}$  is the MLE of  $\theta$ , then for any function  $\tau(\theta)$ , the MLE of  $\eta = \tau(\theta)$  is  $\hat{\eta}_{ML} =$

**field:** If  $\hat{\theta}_{ML}$  is the MLE of  $\theta$ , then for any function  $\tau(\theta)$ , the MLE of  $\eta = \tau(\theta)$  is  $\hat{\eta}_{ML} = \tau(\hat{\theta}_{ML})$

**NOTE:**

**field:** 300

**field:** Bias

**field:** Let  $\hat{\theta} = T(\mathbf{X})$  be an estimator of  $\theta$ . Then the Bias of  $\hat{\theta}$  as an estimator of  $\theta$  is defined as

$$B_{\theta}(\hat{\theta}) = E_{\theta}(\hat{\theta} - \theta) = E_{\theta}(\hat{\theta}) - \theta$$

That is the difference between the expected value of  $\hat{\theta}$  and  $\theta$ .

An estimator  $\hat{\theta}$  of  $\theta$  is said to be unbiased if  $B_{\theta}(\hat{\theta}) = 0 \quad \forall \theta$

**NOTE:**

**field:** 301

**field:** Mean Squared Error

**field:** Let  $\hat{\theta} = T(\mathbf{X})$  be an estimate of  $\theta$ . Then, the **Mean Squared Error** (MSE) of  $\hat{\theta}$  as an estimator of  $\theta$  is defined as:

$$MSE(\hat{\theta}) = E_{\theta}[(\hat{\theta} - \theta)^2] = V_{\theta}(\hat{\theta}) + [B_{\theta}(\hat{\theta})]^2$$

**NOTE:**

**field:** 302

**field:** Do unbiased estimators always exist?

**field:** No, Suppose that  $X \sim \text{Binomial}(n, p)$  and let  $\theta = 1/p$  be the parameter of interest. Can we find an unbiased estimator for  $\theta$ ?- No

**NOTE:**

**field:** 303

**field:** UMVUE

**field:** An estimator  $W^*$  is called a best unbiased estimator of  $\tau(\theta)$  if it satisfies  $E_{\theta}(W^*) = \tau(\theta)$ , for all  $\theta$ , and for any other estimator  $W$  with  $E_{\theta}(W) = \tau(\theta)$ , we have  $V_{\theta}(W^*) \leq V_{\theta}(W), \forall \theta$ . Equivalently  $W^*$  is also called a **Uniform Minimal Variance Unbiased Estimator** (UMVUE) of  $\tau(\theta)$

**NOTE:**

**field:** 304

**field:** Finding a UMVUE

**field:** Start with a complete statistic, (find min suff statistic, prove completeness), Find bias (ie  $E(T(\mathbf{X}))$ ). Then adjust  $T(\mathbf{X})$  to be unbiased. (ie center or scale )

**NOTE:**



**field:** 305

**field:** Cramer-Rao Inequality

**field:** Let  $X_1, \dots, X_n$  be a sample with joint pdf or pmf  $f(\mathbf{x}|\theta)$  and let  $W(\mathbf{X}) = W(X_1, \dots, X_n)$  be any estimator satisfying

$$\frac{d}{d\theta} E_\theta(W(\mathbf{X})) = \int \frac{d}{d\theta} [W(\mathbf{X}) f(\mathbf{x}|\theta)] d\mathbf{x}$$

and  $V_\theta(W(\mathbf{X})) < \infty$

Then,

$$V_\theta(W(\mathbf{X})) \geq \frac{(\frac{d}{d\theta} E_\theta(W(\mathbf{X})))^2}{E_\theta[(\frac{\partial}{\partial \theta} \log f(\mathbf{x}|\theta))^2]}$$

Observe that if the sample  $X_1, \dots, X_n$  is iid with common pdf or pmf  $f(x|\theta)$ , we obtain

$$V_\theta(W(\mathbf{X})) \geq \frac{[\frac{d}{d\theta} E_\theta(W(\mathbf{X}))]^2}{n E_\theta[(\log f(\mathbf{x}|\theta))^2]}$$

The denominator is the information in the sample about  $\theta$

We have that as the information number gets bigger we have a smaller bound for the variance. of the best unbiased estimator and therefore more information is available.

**NOTE:**

**field:** 306

**field:** Cramer-Rao and UMVUE example UMVUE of  $\lambda$  for Poisson

**field:** Poisson example, we have  $\tau(\lambda) = \lambda$ , so  $\frac{d}{d\lambda} \tau(\lambda) = 1$

On the other hand,

$$\begin{aligned}
nE_\lambda\left[\left(\frac{d}{d\lambda}\log f(x|\lambda)\right)^2\right] &= -nE_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda)\right) \\
&= -nE_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)\right) \\
&= -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}(-\lambda + x\log\lambda - \log(x!))\right] \\
&= -nE_\lambda\left(\frac{-x}{\lambda^2}\right) \\
&= \frac{n}{\lambda}
\end{aligned}$$

Therefore, for any unbiased estimator  $W$  of  $\lambda$ , we must have  $V_\lambda(W) \geq \lambda/n$ . Since  $V_\lambda(\bar{X}) = \frac{\lambda}{n}$ , we have that  $\bar{X}$  is an UMVUE of  $\lambda$

**NOTE:**

**field:** 307

**field:** Does  $S^t$  for Normal attain cramer rao?

**field:** No - Suppose that  $X_1, \dots, X_n$  are iid  $N(\mu, \sigma^2)$  and consider the estimation of  $\sigma^2$  when  $\mu$  is unknown.

We have that

$$\frac{\partial^2}{\partial(\sigma^2)^2} \log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\right] = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

and

$$\begin{aligned}
-E\left[\frac{\partial^2}{\partial(\sigma^2)^2} \log f(x|\mu, \sigma^2)\right] &= -E\left(\frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}\right) \\
&= -\frac{1}{2\sigma^4} + \frac{\sigma^2}{\sigma^6} \\
&= \frac{1}{2\sigma^4}
\end{aligned}$$

and therefore, any unbiased estimator  $W$  of  $\sigma^2$  must satisfy  $V(W) \geq \frac{2\sigma^4}{n}$ . Recall that for  $S^2$  we have

$$V(S^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$$

and therefore  $S^2$  does not attain the cramer-rao lower bound.

**NOTE:**

**field:** 308

**field:** Rao-Blackwell

**field:** Let  $W$  be any unbiased estimator  $\tau(\theta)$  and let  $T$  be a sufficient statistic for  $\theta$ . Define  $\phi(T) = E(W|T)$ . Then  $E_\theta(\phi(T)) = \tau(\theta)$  and  $V_\theta(\phi(T)) \leq V_\theta(W)$ , for all  $\theta$ . That is,  $\phi(T)$  is a uniformly better unbiased estimator of  $\tau(\theta)$

**NOTE:**

**field:** 309

**field:** Use of Rao-Blackwell

**field:** Estimators can be improved (their MSE) using sufficiency (already sufficient statistics, or functions of sufficient statistics cannot be improved)

**NOTE:**

**field:** 310

**field:** Are unbiased estimators based on complete sufficient statistics unique.

**field:** Unbiased estimators based on complete sufficient statistics are unique.

**tags:** Theory3

**NOTE:**

**field:** <sub>311</sub>

**field:** Data summaries vs Prediction vs Inference

**field:**

- Data summaries: descriptive statistics summarizing a dataset (ie sample mean)
- Prediction: Use patterns in a data-set to make predictions regarding values of new observations
  - Prediction setting is more flexible than inference setting, as we are not trying to make probabilistic inference, assumptions only matter if they affect prediction quality.
- Inference: Use observations in data set to infer information concerning population parameters

**NOTE:**

**field:** <sub>312</sub>

**field:** Parametric Inference

**field:** Inference (estimation and/or hypothesis testing performed under the assumption that the data come from a population distribution that belongs to some family of distributions  $F(x; \theta)$ ) parametrized by a finite-dimensional parameter  $\theta$

Parameter space: The set  $\Theta$  of all possible values of the parameter  $\theta$

Vs Nonparametric Inference - where no or limited assumptions or specifications of the form of the population distributions

**NOTE:**

**field:** 313

**field:** Are the following tests parametric, semiparametric, or nonparametric

- F-test
- Exact binomial test
- Fisher's exact test
- t-test
- Wilcoxon rank sum
- Permutation tests
- Sign test
- Mood's test
- KS test
- t-test

**field:**

- F-test - Parametric
- Exact binomial test
- Fisher's exact test
- t-test
- Wilcoxon rank sum: semiparametric
- Permutation tests
- Sign test: nonparametric
- Mood's test
- KS test
- t-test

**NOTE:**

**field:** 314

**field:** Definition: Simple hypothesis, composite hypothesis

**field:**

- Simple hypothesis: Completely specifies the parameter value and therefore the population distribution. Simple hypothesis have the form  $H_0 : \theta = \theta_0$  and  $H_1 : \theta = \theta_1$ , for specified values of  $\theta_0$  and  $\theta_1$
- Composite hypothesis: Includes more than one possible parameter value. Composite hypotheses have the form  $H_0 : \theta \in \Theta_0$  and  $H_1 : \theta \in \Theta_1$

**NOTE:**

**field:** 315

**field:** Test procedure:

**field:**

- Random Sample (data):  $X_1, \dots, X_n$
- Sample Space  $\mathcal{X}$  the set of all possible observed samples  $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$
- Hypothesis  $H_0 : \theta \in \Theta_0$  and  $H_1 : \theta \in \Theta_1$  with  $\Theta_0 \cap \Theta_1 = \emptyset$
- Rejection Region  $\mathcal{R} \subset \mathcal{X}$ :
  - If  $(X_1, \dots, X_n \in \mathcal{R}, \text{Reject } H_0)$
  - If  $X_1, \dots, X_n \notin \mathcal{R}, \text{Fail to reject } H_0$

Equivalently

- Random Sample (data):  $X_1, \dots, X_n$

- Test statistic  $T(X_1, \dots, X_n)$  is some function of the data, which is itself a random variable
- Test Statistic Sample Space  $\mathcal{T}$  the set of all possible observed samples  $T = t$
- Rejection Region  $\mathcal{R}_T \subset \mathcal{T}$ :
  - If  $T(X_1, \dots, X_n) \in \mathcal{R}_t$ , Reject  $H_0$  )
  - If  $T(X_1, \dots, X_n) \notin \mathcal{R}_t$ , Fail to reject  $H_0$

**NOTE:**

**field:** <sub>316</sub>

**field:** Power function (definition)

**field:** We can summarize the performance of a test procedure through the power function:

$$\begin{aligned}
 \text{Power}(\theta) &= \beta(\theta) = P_\theta(\text{Reject } H_0 \text{ when } \theta \text{ is the true value of the parameter of interest}) \\
 &= P_\theta((X_1, \dots, X_n) \in \mathcal{R}) \\
 &= P_\theta(T(X_1, \dots, X_n) \in \mathcal{R}_t)
 \end{aligned}$$

Equivalently, for a critical function  $\psi$ ,

$$\beta(\theta) = E_\theta(\psi(x_1, \dots, x_n))$$

**NOTE:**

**field:** <sub>317</sub>

**field:** Calculating Type I and Type II errors from the power function

**field:**

$$P(\text{Type I Error when } \theta = \theta_0 \in \Theta_0) = \beta(\theta_0)$$

$$P(\text{Type II Error when } \theta = \theta_1 \in \Theta_1) = 1 - \beta(\theta_1)$$

(Note these are for simple hypotheses), for complex hypothesis, we want to look at the maximum possible error

To work out these probabilities, we need to know the distribution of the test statistics under the null (For type I error) and alternative (For type II error )

**NOTE:**

**field:** 318

**field:** Size of a test procedure

**field:** The size of a test procedure for a null hypothesis  $H_0 : \theta \in \Theta_0$  is the value

$$\sup_{\theta \in \Theta_0} P_{\theta}(\text{Reject } H_0) = \sup_{\theta \in \Theta_0} \beta(\theta)$$

That is, the size of a test procedure is the largest value of the probability of a Type I Error, across all values of  $\theta$  in the null hypothesis set  $\Theta_0$

**NOTE:**

**field:** 319

**field:** Definition of a level  $\alpha$  test

**field:** A hypothesis test procedure is said to be a level  $\alpha$  test if

$$\sup_{\theta \in \Theta_0} P_{\theta}(\text{Reject } H_0) = \sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$$

That is if the size of the test is less than or equal to  $\alpha$ , the test is a level  $\alpha$  test.

**NOTE:**



**field:** 320

**field:** most powerful level  $\alpha$  test (definition)

**field:** Let  $\mathcal{C}_\alpha$  be the set of all tests of  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$  where  $\Theta_0 \cap \Theta_1 = \emptyset$  that have level  $\alpha$ . A test belonging to  $\mathcal{C}_\alpha$  is the most powerful level  $\alpha$  test at  $\theta_1 \in \Theta_1$  if

$$\beta(\theta_1) \geq \beta^*(\theta_1)$$

for any other test in  $\mathcal{C}_\alpha$  with power function  $\beta^*(\theta)$

**NOTE:**

**field:** 321

**field:** Uniformly most powerful level  $\alpha$  test (definition )

**field:** A test belonging to  $\mathcal{C}_\alpha$  with power function  $\beta(\theta)$  is uniformly most powerful level  $\alpha$  if it is the most powerful for every  $\theta_1 \in \Theta_1$

**NOTE:**

**field:** 322

**field:** Critical Function / Test Function (definition )

**field:** A function  $\psi : \mathcal{X} \rightarrow [0, 1]$  such that  $\psi(x_1, \dots, x_n)$  is the probability of rejecting  $H_0$  when the sample  $(X_1 = x_1, \dots, X_n = x_n)$  is observed is called a critical function of a test procedure.

**NOTE:**

**field:** 323

**field:** Randomized Test (definition)

**field:** A test procedure with critical function  $\psi$  for which there are some points in the sample space such that  $0 < \psi < 1$  is called a randomized test (often used in discrete cases)

**NOTE:**

**field:** 324

**field:** Finding the most powerful level  $\alpha$  test of a simple null hypothesis vs a simple alternative hypothesis

**field:** (*Neyman – Pearson*) The most powerful level  $\alpha$  test of a simple null hypothesis  $H_0$  vs a simple alternative hypothesis  $H_1$  based on data  $\mathbf{X}$  is given by the critical function

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } \frac{L(H_0:x)}{L(H_1:x)} < k \\ c & \text{if } \frac{L(H_0:x)}{L(H_1:x)} = k \\ 0 & \text{if } \frac{L(H_0:x)}{L(H_1:x)} > k \end{cases}$$

Where the constants  $k$  and  $c$  are chosen to ensure that  $E_{H_0}(\psi(\mathbf{X})) = \alpha$

**NOTE:**

**field:** 325

**field:** Steps for using Neyman-Pearson to obtain most powerful tests for simple and alternative hypotheses:

**field:**

1. Identify the likelihood under the null  $L(H_0 : x)$  and alternative  $L(H_1 : x)$
2. Determine how the ratio of the likelihoods  $\frac{L(H_0:x)}{L(H_1:x)}$  depends on the observed data  $\mathbf{x}$  (ie is it an increasing or decreasing function of  $T(\mathbf{X})$ )?
3. Identify the null distribution of the statistic  $T(\mathbf{X})$

- (a) If  $\frac{L(H_0:x)}{L(H_1:x)}$  is an increasing function of  $T(\mathbf{x})$ , rejecting for small values of  $\frac{L(H_0:x)}{L(H_1:x)}$  is equivalent to rejecting for small values of  $T(\mathbf{x})$ , so find  $k$  such that

$$P_{H_0}(T(\mathbf{x}) < k) = \alpha$$

- (b) If  $\frac{L(H_0:x)}{L(H_1:x)}$  is a decreasing function of  $T(\mathbf{x})$ , rejecting for large values of  $\frac{L(H_0:x)}{L(H_1:x)}$  is equivalent to rejecting for large values of  $T(\mathbf{x})$ , so find  $k$  such that

$$P_{H_0}(T(\mathbf{x}) > k) = \alpha$$

**NOTE:**

**field:** 326

**field:** Uniformly most powerful (UMP) level  $\alpha$  test procedure

**field:** Uniformly most powerful (UMP) level  $\alpha$  test procedure for testing  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$  is one with power function  $\beta(\theta)$  such that for every  $\theta_1 \in \Theta_1$  we have

$$\beta(\theta) \geq \beta^*(\theta)$$

for any other level  $\alpha$  test procedure with power function  $\beta^*(\theta)$

**NOTE:**

**field:** 327

**field:** Monotone likelihood ratio

**field:** The family of distributions  $\{F(x|\theta)\}$  indexed by parameter  $\theta \in \Theta$  has monotone likelihood ratio if there is a statistic  $T(\mathbf{X})$  such that for all  $\theta^* > \theta \in \Theta$  and  $\mathbf{x} \in \mathcal{X}$ , the likelihood ratio

$$\frac{L(\theta^*|\mathbf{x})}{L(\theta|\mathbf{x})} \quad \text{is monotone nondecreasing in } T(\mathbf{x})$$

**NOTE:**

**field:** 328

**field:** How to find the UMP test of a simple null hypothesis vs a one sided complex alternative

**field:** See if the family has monotone likelihood ratio in  $T(\mathbf{x})$  UMP tests of one sided alternative hypothesis exist and are given by the form in Neyman-Pearson (by Karlin Rubin )

**NOTE:**

**field:** 329

**field:** Karlin-Rubin Theorem

**field:** Suppose the family of distributions  $\{F(x|\theta)\}$  indexed by parameter  $\theta \in \Theta$  has monotone likelihood ratio Then to test

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

the test function

$$\phi(\mathbf{X}) = \begin{cases} 1 & \text{if } T(\mathbf{X}) > k \\ \gamma & \text{if } T(\mathbf{X}) = k \\ 0 & \text{if } T(\mathbf{X}) < k \end{cases}$$

Where  $k$  and  $\gamma$  are chosen so that  $E_{\theta_0}(\phi(\mathbf{X})) = \alpha$  gives a uniformly most powerful (UMP) level  $\alpha$  test.

(note if we have a one sided lower alternative, we flip the direction of the inequalities )

**NOTE:**

**field:** 330

**field:** Is there a UMP two sided test for  $X_1, \dots, X_n$  iid  $\text{Exp}(\lambda)$  where  $H_0 : \lambda = 2$  vs  $H_1 \lambda \neq 2$ ?

**field:** No: For  $\lambda_1 > \lambda_0 = 2$ , the UMP test would have the form

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i > k_1 \\ 0 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i < k_1 \end{cases}$$

and for  $\lambda_1 < \lambda_0 = 2$ , the UMP test would have the form

$$\psi(\mathbf{X}) = \begin{cases} 1 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i < k_2 \\ 0 & \text{if } T(\mathbf{x}) = \sum_{i=1}^n x_i > k_2 \end{cases}$$

Since these forms are not the same, there is no UMP test.

**NOTE:**

**field:** <sub>331</sub>

**field:** Let  $X \sim \text{Unif}(0, \theta)$ . Is there a UMP test for testing two sided  $H_0 : \theta = 1$  vs  $H_1 : \theta \neq 1$

**field:** Yes:

$$\psi(\mathbf{x}) = \begin{cases} 1 & x < \alpha \text{ or } x > 1 \\ 0 & \alpha < x < 1 \end{cases}$$

**NOTE:**

**field:** <sub>332</sub>

**field:** Unbiased test (definition)

**field:** A test of  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$  is called unbiased if  $\beta(\theta_1) \geq \beta(\theta_0)$  for all  $\theta_1 \in \Theta_1$  and all  $\theta_0 \in \Theta_0$

**NOTE:**

**field:** 333

**field:** Uniformly most powerful unbiased (UMPU) level  $\alpha$  test (definition)

**field:** A level  $\alpha$  test of  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$  with critical function  $\psi(\mathbf{x})$  is called uniformly most powerful unbiased (UMPU) if it is unbiased level  $\alpha$  and for any other unbiased test with critical function  $\psi^*(\mathbf{x})$ , we have

$$E_\theta(\psi(\mathbf{x})) \geq E_\theta(\psi^*(\mathbf{x})) \quad \text{for all } \theta \in \Theta_1$$

**NOTE:**

**field:** 334

**field:** Likelihood Ratio Test (definition)

**field:** Suppose we have the data  $\mathbf{X} = X_1, \dots, X_n$ , with joint density  $f(x; \theta)$  for some parameter  $\theta \in \Theta$ , and we wish to perform a level  $\alpha$  test of  $H_0 : \theta \in \Theta_0$  vs  $H_1 : \theta \in \Theta_1$ , where  $\Theta_1 \cup \Theta_0 = \Theta$ . The likelihood ratio test statistic is given by

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta|x)}{\sup_{\theta \in \Theta} L(\theta|x)} = \frac{L(\hat{\theta}_{0,MLE}; x)}{L(\hat{\theta}_{MLE}; x)}$$

and the null hypothesis is rejected for small values of  $\lambda$  (indicating that the null hypothesis is relatively 'unlikely')

We maximize by finding  $\theta = \hat{\theta}_{MLE}$  and  $\hat{\theta}_{0,MLE}$

**NOTE:**

**field:** 335

**field:** If  $T(\mathbf{X})$  is a sufficient statistic for  $\theta$ , then  $\lambda(\mathbf{x})$  (the LRT statistic)...

**field:** If  $T(\mathbf{X})$  is a sufficient statistic for  $\theta$ , then  $\lambda(\mathbf{x})$  will be a function of  $T(\mathbf{x})$ . In particular  $\lambda(\mathbf{x})$  will be a function of the minimal sufficient statistic

**NOTE:**

**field:** 336

**field:** Frequentist Probability vs Bayesian probability (definition)

**field:**

- Frequentist: For an event  $E$ , in our outcome space,  $P(E)$  is the long run proportion of experiments that have outcome  $E$ , the relative frequency with which an event happens is its probability
- Bayesian: For an event  $E$  in the outcome space,  $P(E)$  is any number between zero and one that you want to assign it, as long as you are coherent about the rules of additivity etc.

**NOTE:**

**field:** 337

**field:** Treatment of population parameters, frequentist vs bayesian

**field:**

- Frequentist: A population parameter  $\theta$  is some fixed (though generally unknown value) that belongs to some set of possible values  $\Theta$
- Bayesian: A population parameter  $\theta$  is a random quantity that has a prior distribution

**NOTE:**

**field:** 338

**field:** Likelihood function (bayesian )

**field:** Given some value of the parameter  $\theta$ , the distribution of the data  $\mathbf{x}$  is  $f(\mathbf{x}; \theta)$  is the likelihood (a function of both the value  $\theta$  and the data  $\mathbf{x}$ ).

**NOTE:**

**field:** 339

**field:** Posterior Distribution (definition)

**field:** The posterior distribution of theta given the observed data  $\mathbf{x}$  is

$$k(\theta; \mathbf{x}) = \frac{f(\mathbf{x}; \theta)h(\theta)}{\int_{\theta} f(\mathbf{x}; \theta)h(\theta)d\theta}$$

Posterior probability  $\propto$  Likelihood  $\times$  prior probability

Note that the posterior distribution is proportional to the numerator.

**NOTE:**

**field:** 340

**field:** Conjugate Priors

**field:** If the prior  $h(\theta)$  belongs to some (parametric) family of distributions  $\mathcal{P}$  and the likelihood  $L(\theta; \mathbf{x})$  (the joint density of the data for any particular value of  $\theta$ ) is such that the posterior  $k(\theta; \mathbf{x})$  belongs to the same family  $\mathcal{P}$ , then this family of priors is said to be conjugate for the likelihood  $L(\theta; \mathbf{x})$  (ie the posterior family is the prior family if we choose a conjugate prior. )

**NOTE:**

**field:** 341

**field:** Non-informative Priors

**field:** A non-informative prior is intended to give as little information as possible about the value of the parameter of interest  $\theta$ .

**NOTE:**



**field:** <sup>342</sup>

**field:** Improper prior

**field:** an improper prior is a prior that does not integrate to one

**NOTE:**

**field:** <sup>343</sup> ©

**field:** Bayes Estimator

**field:** A Bayes estimator (with respect to the particular prior/likelihood) is the estimator that minimizes the Bayesian Risk

$$\delta^* = \operatorname{arginf}_{\delta \in D} \int_{\Theta} R(\theta, \delta) h(\theta) d\theta$$

Where D is the set of all possible estimators for  $\theta$

The Bayes Estimator equivalently minimizes the posterior risk, given the observed data.

For squared-error loss, the Bayes estimate is the mean of the posterior distribution  $k\theta(\mathbf{x})$ :

$$\delta^* = \int_{\Theta} \theta k(\theta|\mathbf{x}) d\theta$$

**NOTE:**

**field:** <sup>344</sup>

**field:** Maximum A Posteriori (MAP)

**field:** A MAP test selects the hypothesis  $H_0$  or  $H_1$  that has the highest posterior probability. (Bayesian. )

**NOTE:**

**field:** 345

**field:** Definition of a p-value

**field:**

- For testing null hypothesis  $H_0$  vs alternative hypothesis  $H_1$ , the p-value  $p(\mathbf{x})$  corresponding to the observed data, is the smallest value  $\alpha$  for which  $H_0$  would be rejected by a size  $\alpha$  test
- Let  $W(\mathbf{X})$  be a test statistic such that large values of  $W$  are evidence that  $H_1$  is true, and therefore the null hypothesis  $H_0$  is rejected for large  $W(\mathbf{X})$  then a p-value can be defined as

$$p(\mathbf{x}) = \sup_{\theta \in \Theta} P_{\theta}(W(\mathbf{X}) \geq W(\mathbf{x}))$$

This says that the p-value is the (largest in the null space) probability of obtaining a test statistic at least as extreme as the observed test statistic value.

- A p-value is just a function of the observed data; a test statistic

**NOTE:**

**field:** 346

**field:** Validity of p-value

**field:** A p-value is valid (exact) if for every  $\theta \in \Theta_0$  and every value of  $\alpha \in [0, 1]$ , we have

$$P_{\theta}(p(\mathbf{X}) \leq \alpha) \leq \alpha$$

**NOTE:**

**field:** 347

**field:** Confidence interval (definition)

**field:** Suppose we have data  $\mathbf{X}$  such that the (joint) density of our data give information about an unknown parameter  $\theta$ . Then a  $(1 - \alpha)100$  confidence interval for  $\theta$  is a random interval  $[L(\mathbf{X}), U(\mathbf{X})]$  such that

$$\inf_{\theta \in \Theta} P_{\theta}(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X})) = 1 - \alpha$$

It is important to note that it is the limits of the interval  $L(\mathbf{X}), U(\mathbf{X})$  that are the random quantities here.

**NOTE:**

**field:** <sup>348</sup>

**field:** Construct a CI using a hypothesis test

**field:** A level  $(1 - \alpha)100$  confidence interval can be constructed by inverting a level  $\alpha$  hypothesis test. This fact is known as the duality of confidence intervals and hypothesis testing. The confidence region  $\mathcal{C}$   $\mathcal{C} = \{\theta_0 : H_0 : \theta = \theta_0 \text{ would not be rejected at level } \alpha\}$   
(ie solve for  $\theta_0$  to be in the center)

**NOTE:**

**field:** <sup>349</sup>

**field:** Pivot

**field:** Suppose  $X$  comes from some parametric family  $F(\mathbf{x} : \theta)$  indexed by parameter  $\theta$ . A pivot, or pivotal quantity is a random variable  $U = g(\mathbf{X}, \theta)$  that depends upon both the sample  $\mathbf{X}$  and the unknown parameter  $\theta$  for which the distribution of  $U$  does not depend on  $\theta$

**NOTE:**

**field:** <sup>350</sup>

**field:** Finding a confidence interval for  $\theta$  using the pivotal method

**field:**

1. Identify a pivotal quantity  $U$  and its distribution  $F_U(u)$
2. Find  $a$  and  $b$  such that

$$P(a < U < b) = 1 - \alpha$$

Let  $F_U(u)$  denote the cdf of the pivot  $U$ , so then we can set

$$\begin{aligned}a &= F_U^{-1}(c\alpha) \\ b &= F_U^{-1}(1 - (1 - c)\alpha)\end{aligned}$$

For any  $c \in [0, 1]$  (usually .5 to split up area on the tails evenly)

3. Solve the inequality  $a < U < b$  for  $\theta$  in the middle.

**NOTE:**

**field:** <sub>351</sub>

**field:** Pivotal CI example: Let  $Y \sim \exp(\theta)$ .

**field:**

1. Let  $U = Y/\theta$ , so  $U \sim \text{Exp}(1)$ . which doesn't depend on  $\theta$ , so  $U$  is a pivotal quantity.
2. We must find  $a, b$ , such that  $P(a \leq U \leq b) = 1 - \alpha$ . We then solve  $P(U \leq a) = \alpha/2$  and  $P(b \leq U) = \alpha/2$  Solve for  $\theta$

$$P(a \leq U \leq b) = P(Y/b \leq \theta \leq Y/a)$$

**NOTE:**

**field:** <sub>352</sub>

**field:** Finding a pivotal quantity

**field:**

- If  $\theta$  is a location parameter, a possible pivot has the form  $U = T(\mathbf{X}) - a(\theta)$
- If  $\theta$  is a scale parameter,  $U = T(\mathbf{X})/b(\theta)$  is a possible pivot
- If  $\theta$  is a location-scale parameter,  $U = (T(\mathbf{X}) - a(\theta))/b(\theta)$  is a possible pivot
- If neither, use how parameter is related to  $X$ , ie if  $F_Y(y) = y^N$ , use  $y^N$  as the pivot.

**NOTE:**

**field:** <sub>353</sub>

**field:** Confidence Interval optimality criteria

**field:**

- Length
  - Length can be a function of sample size and critical value choice (Normal mean with known variance), other cases a random quantity (Normal mean with unknown variance)
  - When length is random, we typically want confidence intervals with shortest mean length.
  - Ie for normal variance with known mean, the expected length of the confidence interval is  $E(L) = c\sigma^2n$ , so a shortest interval would be not using equal tails, but requires numerical computation.
- Convexity
  - In the case of a one-dimensional parameter, this means that the region should be a connected interval

- Agreement with a reasonable estimate and with a reasonable hypothesis test.
  - Example score test and wald CI
- Equal coverage probability for all  $\theta$

**NOTE:**

**field:** <sub>354</sub>

**field:**  $1 - \alpha$  credible interval

**field:** (Bayesian) A  $1 - \alpha$  credible interval is an interval  $[a, b]$  such that the credible probability of that interval is  $1 - \alpha$ , that is

$$\int_{\theta=a}^b k(\theta|\mathbf{x})d\theta = 1 - \alpha$$

Where the posterior distribution  $k(\theta|\mathbf{x}) = \frac{f(\mathbf{x};\theta)h(\theta)}{\int_{\theta} f(\mathbf{x};\theta)h(\theta)} = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})}$ , where the posterior is proportional to the likelihood times the prior

In many cases, we can choose  $c = .5$  and let  $a =$  the  $c \times \alpha$  quantile of the distribution, and  $b =$  the  $1 - (1 - c) \times \alpha$  quantile of the posterior distribution, but in cases when the posterior distribution is unimodal, we can obtain shorter intervals.

**NOTE:**

**field:** <sub>355</sub>

**field:** Credible probability

**field:** (Bayesian ) For any region  $\mathcal{A} \subset \Theta$ , the credible probability of the set  $\mathcal{A}$  is

$$P(\theta \in \mathcal{A}|\mathbf{x}) = \int_{\theta \in \mathcal{A}} k(\theta|\mathbf{x})d\theta$$

**NOTE:**

**field:** <sub>356</sub>

**field:** Highest posterior density  $1 - \alpha$  credible interval

**field:** For a unimodal posterior distribution  $k(\theta|\mathbf{x})$  Highest posterior density  $1 - \alpha$  credible interval is the interval  $[a, b]$  such that

1.  $\int_{\theta=a}^b k(\theta|\mathbf{x})d\theta = 1 - \alpha$
2.  $k(a|\theta) = k(b|\theta)$
3.  $a < \theta^* < b$  where  $\theta^*$  is the mode of the posterior distribution.

This credible interval will be the shortest. Note it might not contain the Bayes estimator for  $\theta$  if the posterior distribution is very highly skewed and has a heavy tail.

**NOTE:**

**field:** <sub>357</sub>

**field:** Unimodal pdf

**field:** A pdf is unimodal if there exists some value  $x^*$  such that

1. For all  $x \leq x^*$ ,  $f(x)$  is non decreasing
2. For all  $x \geq x^*$ ,  $f(x)$  is non-increasing.

**NOTE:**

**field:** <sub>358</sub>

**field:** Confidence Intervals for functions of parameters: Obtain a confidence interval for  $\tau = g(\theta)$ , given a confidence interval for  $\theta$

**field:**

1. Invert a test of  $H_0 : \tau = \tau_0$  vs  $H_1 : \tau \neq \tau_0$
2. Create a pivot  $U = h(\mathbf{X}, \tau)$ , and construct a pivotal interval for  $\tau$
3. Transform the CI for  $\theta$  into an interval for  $\tau$ 
  - Coverage of this interval:  $C_\tau = \{\tau_0 : \tau_0 = g(\theta_0) \text{ for } \theta_0 \in C_\theta\}$
  - NOT the same as  $(g(L_\theta(\mathbf{X})), g(U_\theta(\mathbf{X})))$  (unless strictly monotone, otherwise coverage not the same. )

**NOTE:**

**field:** 359

**field:** Joint Confidence Intervals for Multivariate parameters

**field:**

1. Construct  $(1 - \alpha^*)$  CIs for each component  $\theta_j$  separately
2. Define  $\mathcal{C} = \{\theta_0 : \theta_{0,j} \in \mathcal{C}_j \text{ for all } j = 1, \dots, p\}$
3. Chose the univariate coverage levels  $(1 - \alpha^*)$  to ensure that the coverage level of the joint confidence region is at least the desired level  $(1 - \alpha)$
4. Often choose  $\alpha^* = \alpha/p$ , although this often gives larger CIs - over-coverage
5. If parameters independent, if we set  $1 - \alpha = (1 - \alpha^{**})^p$

**NOTE:**

**field:** 360

**field:** Example: Construct a CI for

$$X_1, \dots, X_n \sim iidExp(\mu_x, \sigma)$$

$$Y_1, \dots, Y_n \sim iidExp(\mu_x, \sigma)$$



**field:**  $\kappa(\theta) = \mu_x - \mu_y$

- Sufficient statistic for parameter  $(\mu_x, \mu_y, \sigma)$  is

$$(X_{(1)}, Y_{(1)}, W = \sum_{i=1}^n (X_i - X_{(1)}) + \sum_{i=1}^n (Y_i - Y_{(1)}))$$

- $X_{(1)} \sim \exp(\mu_x, \sigma/n)$
- $Y_{(1)} \sim \exp(\mu_y, \sigma/n)$
- $\frac{2W}{\sigma} \sim \chi_{4n-4}^2$
- Construct pivot:  $U = \frac{|n(X_{(1)} - Y_{(1)} - \mu_x + \mu_y)|}{W/(2n-2)} \sim F_{2, 4n-4}$

**NOTE:**

**field:** <sub>361</sub>

**field:** Determinant of a square matrix

**field:** The determinant of a square matrix  $\mathbf{A}$  is denoted by  $|\mathbf{A}|$  and is defined recursively

- For a  $(1 \times 1)$  matrix  $|\mathbf{A}| = a_{11}$
- For a  $(p \times p)$  matrix  $|\mathbf{A}| = \sum_{j=1}^p a_{ij}(-1)^{i+j}|\mathbf{A}_{-i, -j}|$
- Where  $\mathbf{A}_{-i, -j}$  is the matrix obtained by removing the  $i$ th column and  $j$ th row from  $\mathbf{A}$

**NOTE:**

**field:** <sub>362</sub>

**field:** Properties of the determinant

- $\det(\mathbf{I}_p)$
- $\det(\mathbf{A}^{-1})$
- $\mathbf{A}$  is invertible iff
- $\det(AB)$
- $\det(\mathbf{A}^t)$
- $\det(\mathbf{XAX}^{-1}) =$
- Relationship to eigenvalues

**field:**

- $\det(\mathbf{I}_p) = 1$
- $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$
- $\mathbf{A}$  is invertible iff  $\det(\mathbf{A}) \neq 0$
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- $\det(\mathbf{A}^t) = \det(\mathbf{A})$
- $\det(\mathbf{XAX}^{-1}) = \det(\mathbf{A})$
- The determinant of  $\mathbf{A}$  is equal to the product of all eigen values of  $\mathbf{A}$

**NOTE:**

**field:** <sub>363</sub>

**field:** Properties of the Multivariate Normal Distribution

**field:**

- If  $\mathbf{X}$  has a multivariate normal distribution, then each element has a marginal normal distribution
- Random variables  $X_1, \dots, X_p$  with marginal normal distributions DO NOT necessarily have a multivariate normal joint distribution
- All subsets of elements of  $\mathbf{X}$  have a multivariate normal distribution
- All linear combinations of the components of  $X$  are normally distributed
- $\mathbf{X} + c \sim MVN(\mu + c, \Sigma)$
- $\mathbf{A}\mathbf{X} \sim MVN(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^t)$ , where each element of  $\mathbf{A}\mathbf{X}$  is a linear combination of the random vector  $\mathbf{X}$
- $Cov(X_j, X_k) = \sigma_{jk} = 0$  iff and only if  $X_j, X_k$  independent
- $Z = \Sigma^{-1/2}(\vec{X} - \vec{\mu}) \sim MVN(\vec{0}, I_p)$
- $Z^t Z \sim \chi_p^2$

**NOTE:**

**field:** <sub>364</sub>

**field:** Distribution of sample mean and variance from Multivariate normal

**field:**

- $\bar{\mathbf{X}} \sim MVN(\vec{\mu}, \frac{1}{n}\Sigma)$
- $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t$

**NOTE:**

**field:** <sub>365</sub>

**field:**  $X \sim \chi_{v_1}^2, Y \sim \chi_{v_2}^2, X + Y \sim$

**field:**  $\chi_{v_1+v_2}^2$

**NOTE:**

**field:** 366

**field:** Definition of consistent sequence of estimators

**field:** Suppose we have random variables  $X_1, X_2, \dots$ , such that the collection  $\{X_1, \dots, X_n\}$  gives information about a parameter  $\theta$ . A sequence of estimators  $W_n = W_n(X_1, \dots, X_n)$  is a consistent sequence of estimators of the parameter  $\theta$  if FOR EVERY  $\theta \in \Theta$ ,

$$W_n \xrightarrow{p} \theta$$

**NOTE:**

**field:** 367

**field:** Mean square error

**field:**

$$MSE(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2] = Bias^2 + V(\hat{\theta})$$

**NOTE:**

**field:** 368

**field:** If a sequence of estimators  $W_n$  for a parameter  $\theta$  satisfies

- $\lim_{n \rightarrow \infty} V_\theta(W_n) \rightarrow 0$
- $\lim_{n \rightarrow \infty} E_\theta(W_n) \rightarrow \theta$
- Equivalently, if  $MSE_\theta(W_n; \theta) \rightarrow 0$

for all  $\theta \in \Theta$ , then

**field:**  $W_n$  is a consistent sequence of estimators for the parameter  $\theta$

**NOTE:**

**field:** 369

**field:** How to prove consistency:

**field:** If a sequence of estimators  $W_n$  for a parameter  $\theta$  satisfies

- $\lim_{n \rightarrow \infty} V_\theta(W_n) \rightarrow 0$
- $\lim_{n \rightarrow \infty} E_\theta(W_n) \rightarrow \theta$
- Equivalently, if  $MSE_\theta(W_n; \theta) \rightarrow 0$

for all  $\theta \in \Theta$ , then  $W_n$  is consistent.

**NOTE:**

**field:** 370x

**field:** Definition of asymptotic distribution of Estimators

**field:** Suppose that a sequence of random variables  $W_n$  satisfies

$$k_n(W_n - \theta) \xrightarrow{d} F$$

- $k_n$  is the stabilizing constant,
- $F$  is the asymptotic distribution
- $\sigma^2(\theta)$  the asymptotic variance (which can't depend on  $n$ )

**NOTE:**

**field:** <sub>371x</sub>

**field:**  $X_i \sim \text{Exp}(\sigma), \sum X_i \sim$

**field:**  $\sum X_i \sim \text{Gamma}(n, \sigma)$

**NOTE:**

**field:** <sub>372</sub>

**field:** Relationship between asymptotic variance and limiting variance

**field:** limiting variance  $\geq$  asymptotic variance, where the limiting variance is defined as  $\lim_{n \rightarrow \infty} V(Y_n)$

**NOTE:**

**field:** <sub>373</sub>

**field:** Score function

**field:**

$$U(\theta) = \frac{\partial}{\partial \theta} l(\theta; \mathbf{x}) = \sum_{i=1}^n \frac{\frac{\partial}{\partial \theta} f(x_i; \theta)}{f(x_i; \theta)}$$

**NOTE:**

**field:** <sub>374</sub>

**field:** Notation and test construction  $X_1, \dots, X_n$  iid

- Density/Mass Function
- Likelihood
- Log likelihood
- Maximum likelihood estimator
- Score
- Information

**field:**

- Density/Mass Function :  $f(x; \theta)$
- Likelihood :  $L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta)$
- Log likelihood  $l(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}) = \sum_{i=1}^n \log f(x_i; \theta)$
- Maximum likelihood estimator  $\hat{\theta}_n : \frac{\partial}{\partial \theta} l(\theta; \mathbf{x}) \Big|_{\theta=\hat{\theta}_n} = 0$  (except if support depends on parameter of interest)
- Score:  $U(\theta) = \frac{\partial}{\partial \theta} l(\theta; \mathbf{x}) = \sum_{i=1}^n \frac{\frac{\partial}{\partial \theta} f(x_i; \theta)}{f(x_i; \theta)}$
- Information:  $I_1(\theta) = E \left( - \frac{\partial^2}{\partial \theta^2} \log f(x; \theta) \right)$

**NOTE:**

**field:** <sub>375</sub>

**field:** Important Distribution of Maximum Likelihood Estimators

**field:** Note these assumptions hold in exponential families, Binomial, Poisson, etc

- $X_i$  iid with common density function
- Identifiability: for  $\theta_1 \neq \theta_2$ ,  $f(x; \theta_1) \neq f(x; \theta_2)$
- Common support (the one we need to verify ) The set of possible values for  $X$  does not depend on the value of the parameter  $\theta$
- Open Parameter Space (ie cant include 0 or 1 in bernoulli case)

**NOTE:**

**field:** <sup>376</sup>

**field:** Asymptotic Normality of MLE

**field:** Under the common assumptions, The MLE  $\hat{\theta}_n$  for  $\theta$  satisfies

$$\frac{\hat{\theta}_n - \theta}{\sqrt{\frac{1}{nI_1(\theta)}}} \xrightarrow{d} N(0, 1)$$

Equivalently,  $\hat{\theta}_n \sim N(\theta, \frac{1}{nI_1(\theta)})$  This implies that the MLE is asymptotically the UMVUE (as  $E(\hat{\theta}_n) \approx \theta$  and  $Var(\hat{\theta}_n) \approx \frac{1}{nI_1(\theta)}$ ), the CRLB var of unbiased estimate of  $\theta$

**NOTE:**

**field:** <sup>377</sup>

**field:** Mann-Wald Theorem (for MLE)

**field:** For any differentiable  $g(\cdot)$  with non-zero first derivative: if  $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{I_1(\theta)})$ , then

$$\sqrt{n(g(\hat{\theta}_n) - g(\theta))} \xrightarrow{d} N(0, \frac{[g'(\theta)]^2}{I_1(\theta)})$$



**NOTE:**

**field:** 378

**field:** Wald Hypothesis Tests and Confidence intervals

**field:** Since  $\theta$  is not known, we cannot find the value of  $I(\theta)$  exactly, but we can use  $\hat{I}(\theta) = I(\hat{\theta}_n) \xrightarrow{P} I(\theta)$ , where  $\hat{\theta}_n$  is the MLE

- Test statistic:

$$W(\theta_0) = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{1}{nI_1(\hat{\theta}_n)}}}$$

- Two sided alternative: Reject  $H_0 : \theta = \theta_0$  vs  $\theta \neq \theta_0$  if

$$|W(\theta_0)| > z_{\alpha/2}$$

- Confidence interval:

$$\left( \hat{\theta}_n - z_{\alpha/2} \sqrt{\frac{1}{nI_1(\hat{\theta}_n)}}, \left( \hat{\theta}_n + z_{\alpha/2} \sqrt{\frac{1}{nI_1(\hat{\theta}_n)}} \right) \right)$$

**NOTE:**

**field:** 379

**field:** Asymptotic Likelihood-Based Test Statistics

	Test	Test Statistic	Asymptotic Null Distribution
	Likelihood Ratio	$G(\theta_0) = 2(l(\hat{\theta}_n; \mathbf{x}) - l(\theta_0; \mathbf{x}))$	$\chi_1^2$
<b>field:</b>	Wald	$W(\theta_0) = \frac{(\hat{\theta}_n - \theta_0)^2}{\frac{1}{nI_1(\hat{\theta}_n)}}$	$\chi_1^2$
		$W^*(\theta_0) = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{1}{nI_1(\hat{\theta}_n)}}}$	
	Score	$S(\theta_0) = \frac{U(\theta_0)^2}{nI(\theta_0)}$ (no need to find MLE) $S^*(\theta_0) = \frac{U(\theta_0)}{\sqrt{nI(\theta_0)}}$	$\chi_1^2$ $N(0, 1)$

**NOTE:**

**field:** 380

**field:** Comparison between Likelihood Ratio, Wald and Score tests and intervals

**field:**

1. Likelihood Ratio:

- Requires computation of  $\sup_{\theta \in \Theta} L(\hat{\theta}; \mathbf{x})$  (likelihood of MLE) and  $L(\theta_0; \mathbf{x})$  (likelihood of null hypothesis )
- Generally hardest to invert to form confidence intervals

2. Wald:

- Only requires computation of MLE and Information under MLE
- Tends to differ from the Score and LR tests more than they do from each other
- Easiest to invert to form confidence intervals

3. Score:

- Only requires computation of Information under Null
- Has best power of the three tests for alternatives close to null
- Typically similar in form to Wald, except variance is estimated under null instead of using MLE

All three tests are asymptotically equivalent under the null hypothesis when  $n$  is large. If the alternative is true, they may differ substantially, no matter the value of  $n$

**NOTE:**

**field:** 381

**field:** Variance Stabilizing Transformation

**field:** To potentially improve confidence interval coverage in settings where there is a mean-variance relationship, we can try to find a function  $g(\cdot)$  such that

$$(g'(\theta))^2 V(\hat{\theta}) = 1$$

, That is the asymptotic variance of  $g(\hat{\theta})$  does not depend on the value of  $\theta$ . This function  $g(\cdot)$  is called the variance stabilizing transformation for the estimate  $\hat{\theta}$